In pursuit of a novel UTI treatment strategy – an *in silico* study of the FimH adhesin

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universtität Basel

von

Adam Zalewski
aus Gdańsk, Polen

Referent: Prof. Dr. Angelo Vedani
Korreferent: Dr. Julie Bouckaert

Basel, 2013

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

Auf Antrag von:

Prof. Dr. A. Vedani, Institut für Molekulare Pharmazie, Universität Basel

Dr. J. Bouckaert, UFR de Biologie, Université des Sciences et Technologies de Lille 1, France

Basel, October 15, 2013

Prof. Dr. Jörg Schibler
Dekan

"I may not have gone where I intended to go,
but I think I have ended up where I needed to be."

Douglas Adams

Acknowledgments

First and foremost, I would like to express my heartfelt gratitude to Prof. Angelo Vedani for the opportunity, supervision, and patience that enabled me to conclude my PhD. I feel that the last three and a half years have made me both a better person and scientist, neither of which would have been possible without his support.

I would also like to thank Prof. Beat Ernst for providing both a scientific challenge and the assistance I required to face it. I have no doubt that participating in the FimH project has made my PhD all the more valuable.

A special "thank you" goes to Dr. Sameh Eid, for being an invaluable friend and support through my PhD. It takes a very special kind of person to withstand me for over three years, let alone to keep on helping me for that long.

My kindest thanks go to Dr. Julie Bouckaert for her enormous contributions towards the subject I worked on and for agreeing to be the co-referee of this thesis. I wouldn't prefer anyone else for this role.

Computational science isn't always very exciting and my time at the Pharmacenter would have been far less enjoyable without my fellow "modelers". Special thanks go to Dr. Martin Smieško, Dr. Gianluca Rossato, and Christoph Sager for all the on- and off-topic discussions, during and after working hours.

Scientific challenges aside, what made my time with the IMP special were the people that constituted it. To name just a few, I would like to thank Jacqueline, Simon, Mirko, Meike, Giulio, Wojtek, Kathi, and Matthias for always being the great people that they are.

Last but not least, I want to thank my family. What I owe you goes far beyond this PhD and I only hope that I can someday repay at least a small portion of it.

Abstract

Carbohydrate-binding proteins (particularly lectins) are frequent targets of present pharmaceutical research. While of high priority due to their involvement in an array of pathophysiological events, these systems (along with their saccharidic ligands) present many challenges related, in part, to the limited means of their in-depth structural exploration. A foothold towards improving the design throughput of carbohydrate-based drugs is provided by the state-of-the-art *in silico* technologies, offering both means of structure generation and refinement as well as rapid screening of putative ligands. In this account, the application of some of these methods (docking, molecular dynamics, and MM-GBSA/QSAR protocols), toward inhibiting FimH – a virulence factor of urinary tract infections – is presented. The role of this bacterial lectin involves the adhesion to highly-mannosylated urothelial cell surfaces. This process is supported by the presence of urine flow, leading to the so-called catch-bonding characteristics. The presented work was focused primarily on the rationalization and optimization of the binding of mannosidic ligands with available crystallographic and *in vitro* data. As a result, several series of new promising compounds were designed, with selected among them synthesized, assayed, and published. In addition, an array of molecular dynamics simulation techniques was employed to probe the structural properties of the native, two-domain protein in presence and absence of shear conditions. Through this, a receptor conformation was identified, that differed (including the binding site) from ones previously employed in structure-based design approaches. Given the wealth of data reporting the existence of FimH variants with a broad range of affinities towards their natural ligands (further dependent on many endogenic and exogenic factors), the reported discoveries may lead to versatile and potent inhibitors, superior to currently applied antibiotic treatments.

The author's contribution

FimH inhibitor design and optimization

All underlying computational work and its interpretation. In addition, I participated in the design of the presented compounds and the writing of two manuscripts awaiting publication (discussed in Subsections "Mutation Study" and "Replacement of a conserved water molecule").

The *VirtualDesignLab*

Preparation of the structural templates, testing and calibration of the underlying protocols. Development of the QSAR model and the entropy calculation module. Co-authorship of the related article.

Modeling on the complete FimH protein

The described work was planned, carried out, and analyzed entirely by me.

Contents

Nomenclature

AMD	Accelerated Molecular Dynamics
FEP	Free Energy Perturbation
IFD	Induced Fit Docking
ITC	Isothermal Titration Calorimetry
LD	Lectin Domain
MD	Molecular Dynamics
MM/GBSA	Molecular Mechanics/Generalized Born Surface Area
MM/PBSA	Molecular Mechanics/Poisson-Boltzmann Surface Area
mQSAR	multidimensional Quantitative Structure-Activity Relationships
NMA	Normal Mode Analysis
PD	Pilin Domain
PDB	Protein Data Bank
PMF	Potential of Mean Force
RMSD	Root Mean Square Deviation
RRHO	Rigid-Rotor Harmonic-Oscillator
SMD	Steered Molecular Dynamics
TI	Thermodynamic Integration
UPEC	Uropathogenic Escherichia Coli
UTI	Urinary Tract Infections

1. Aim of this work

The aim of this work is twofold. Firstly, it describes the support provided – by means of several *in silico* techniques – to the design and study of small molecular ligands aimed at inhibiting the FimH adhesin. This part focuses on predicting and scoring the binding poses of prospective synthesis candidates through flexible docking and molecular dynamics simulations. As my work progressed, some of these tasks were eventually re-delegated to a fully automated platform referred to as the *VirtualDesignLab*, offering superior predictive power and accessibility to all researches involved in the project. A final addition to the platform – that, given the potential of the underlying *VirtualToxLab* technology, enables re-purposing toward other systems of interest – is a module calculating the entropy change sustained by a given system upon ligand binding.

The second part of the thesis describes the attempts of elucidating the dynamic structural properties of the full, two-domain FimH. The significance of this work lies in the fact that the protein is reported to assume multiple conformational states depending on various conditions and resulting in a broad spectrum of ligand binding affinities. In light of this, the determination of whether the available crystal structures are an adequate physiological representation of the system, is of utmost importance. These attempts were based primarily on steered and accelerated molecular dynamics techniques, geared toward exploring structural events inaccessible to the conventional protocols.

2. Introduction

2.1 Properties and biological significance of host-guest complexes

The formation of a complex between two or more molecular entities lies at the heart of nearly every biochemical process. Be it enzymatic cleavage of a substrate, allosteric regulation of a G protein-coupled receptor, or inhibition of a virulence factor, numerous events taking place in living organisms rely on mutual recognition between partner molecules. Among these, the very specific binding of a small molecular ligand to a protein receptor, plays a particularly special role, as already suggested more than a century ago by the German Nobel Prize laureate Emil Fisher. Through his revolutionary work,[1] that would guide scientists in decades to come (paving way for the structure-based design we know today), he presented a model in which a receptor and its ligand fitted each other like the proverbial "lock and key". With this analogy in mind, researches could begin attempts of manufacturing customized "keys" that could stimulate or inhibit the functions of biological receptors involved in various physiological or pathophysiological processes. Eventually, thanks to the advancements in X-ray crystallography, these attempts could be transfered away from an imaginary or two-dimensional space, obtaining insight into the three-dimensional structures of the macromolecular "locks".

Though of great value, the concept of a rigid, perfectly complementary pair of molecules could not explain certain aspects of biochemical recognition. For instance, the binding of molecules smaller than their receptor site as well as promiscuity of certain targets toward a vast array of structurally diverse ligands could not be clarified.[2] It was not until over half a century after the introduction of Fisher's theory, that it could be expanded – largely due to the progress in the fields of crystallography and NMR spectroscopy – with the concepts of flexibility. Henceforth, both the receptor and ligand would no longer be treated as rigid, but rather capable – to an extent – of adapting their shape toward the partner. As a consequence, the design of small

molecules began focusing not only on the conformation in which they would bind to a given protein, but also on how much it deviated from the native conformation in solution. Through this, the concepts of internal strain (i.e. the cost of assuming the bioactive conformation by the ligand) and synthetic structure preorganization (in order to avoid that very strain) could first be introduced into the drug design and optimization attempts.[3,4] Nevertheless, conformational properties of prospective macromolecular targets remained more elusive. The first – and still broadly accepted – model explaining the adaptation of a protein structure toward a small molecular ligand was proposed in the form of the induced-fit (conformational induction) theory.[5] In the context of this model, it is only after the formation of an initial, weak protein-ligand complex, that the former can be "induced" toward forming a shape better accommodating the small molecule. While widely accepted and convenient toward explaining the capability of receptors to adapt various conformations depending on the given ligand, the induction theory is however difficult to validate in context of large – at times spanning multiple protein domains – conformational changes. Also, based on stark differences between the binding site regions of apo and holo receptors,[6] the obligatory initial ligand recognition – requiring a certain degree of complementarity between native protein and ligand conformations – is often used to undermine the induced-fit hypothesis. Hence, alternatively to this mechanism, a conformational selection (population shift, selected fit) model was derived.[7] This theory (ever since gaining increasing attention) postulates, that a vast array of molecular receptors exist within an ensemble of conformational states, separated by small energetical barriers. Between these an equilibrium is proposed, in which only a single, distinct conformation can be "selected" by a binding partner.

With experimental evidence provided to support both receptor adaptation theories, distinctions between them had to be established. Hence, in order to enable categorization of protein-ligand systems between the two mechanisms, a simplified representation of their underlying energy landscapes (referred as the double-basin Hamiltonian) was developed [Figure 2.1].[8] According to the concept, induced-fit events are accompanied by relatively small conformational changes, commonly limited to the side chains of residues directly involved in binding. Thanks to the energetical benefit of initial contact with the ligand, these events can transgress significant energy barriers, often leading to very stable final complex. On the other hand, conformational selection preference is believed to be a trademark of systems with low (several kT at most) transition barriers between the involved states.[9] The strength of ligand binding necessary to shift the underlying equilibrium toward the binding-capable species is commonly comparably low. With the native pro-

tein motion often involving the backbones of entire domains, conformational selection events may however be associated with much greater structural rearrangements than induced-fit.[10] Presently both modes are widely accepted, with a vigorous discussion continuing with regard to their validity and boundaries.

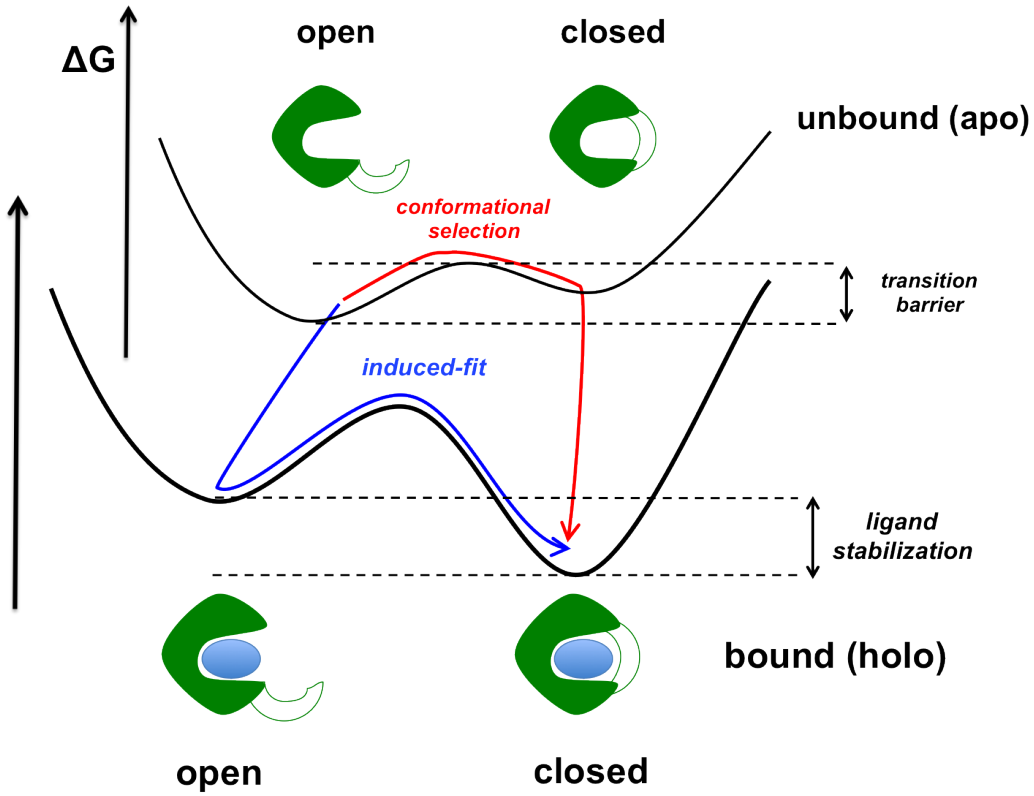


Figure 2.1: Schematic representation of the induced-fit (blue) and conformational selection (red) mechanisms in relation to the protein energy landscape, adapted from the work introducing the double-basin Hamiltonian concept.[8] Highlighted are four distinct conformational states: unbound-open, unbound-closed, bound-open, and bound-closed. Additional vertical arrows correspond to the key energy thresholds likely to determine the preference of one mechanism over the other.

2.2 Thermodynamics of host-guest binding

Although illustrated with simplistic models (e.g. "lock and key", induced-fit, or conformational selection), all binding events relate to the complex theorems of equilibrium thermodynamics. These state that a spontaneous formation of a protein-ligand complex

$$P_{aq.} + L_{aq.} \rightleftharpoons P'L'_{aq.} \tag{2.1}$$

may only take place if the accompanying Gibb's free energy change within the entire system (commonly referred to as the binding affinity) is negative.[11] This change is most often described through the equation

$$\Delta G = -RTlnK_b \tag{2.2}$$

where R and K_b are the universal gas and binding constants and T is the temperature. Reversible in nature and governed by non-bonded interactions, ligand-target binding commonly falls into a K_b range of $10^{-2} - 10^{-12}$ M (sub-millimolar to picomolar), corresponding to free energy changes between 2 and 15 kcal/mol at 298 K.[12] Common drug-target complexes are found between 5 and 12 kcal/mol, with the remaining parts of the spectrum occupied by weak, naturally-occurring organic complexes and very strong enzyme-inhibitor pairs respectively. With regard to the strength and energetical makeup of a specific binding event, the ΔG can also be dissected into enthalpic (ΔH) and entropic (ΔS) components

$$\Delta G = \Delta H - T\Delta S \tag{2.3}$$

Though strongly correlated, leading to the phenomenon of entropy-enthalpy compensation (discussed later in this chapter),[13] the free energy constituents are commonly further decomposed and approached separately, to enable insight into various factors guiding the protein-ligand binding.

2.2.1 Enthalpy

Enthalpy is the cumulative measure of the thermodynamic energy within a system, reflected by the amount of heat produced or consumed by it. It consists of the internal energies of all solute and solvent molecules (directly related to the conformations they have to assume during binding events), along with all favorable and unfavorable non-bonded interactions between them.[14] While often non-specific and overlapping (making them difficult to target and optimize), some of these energetical components fall into distinct categories relevant from the viewpoint of structural design.

Electrostatic interactions

Electrostatics arise from non-uniform charge distributions common in molecular interfaces. Given their substantial reach (exceeding 10 Å) along with the omnipresence of charges in biomolecules, these forces also encompass virtually all atoms in a system.[15] At shorter ranges, electrostatics lead to strong repulsive (attractive) interactions between entities with the same (opposite) charges, determining the complementarity of polar and ionized groups, successfully utilized in numerous scientific endeavors.[16]

Commonly, electrostatic forces are represented using the coulombic relationship between the point charges of two atoms

$$E_{el.} = \frac{q_1 q_2}{4\pi\varepsilon_0 R} \tag{2.4}$$

where q_1 and q_2 are the partial charges of each atom, R is the distance between them, and ε_0 – the dielectric constant of the surrounding medium (approximately 80 in the case of water but around 2–6 inside buried protein cavities). It also bears mentioning, that electrostatic-based interactions dominate the overall free energy and specificity of carbohydrate-protein binding, due to the abundance of polar hydroxyl groups at the corresponding interfaces.[17]

Hydrogen bonds

While still electrostatic in nature, hydrogen bonds reveal several distinct properties that lead toward treating them as a separate group of interactions. These highly specific forces form on the basis of attraction between a hydrogen atom and a pair of electronegative moieties, referred to as the donor and acceptor. They are also characterized by rigorous geometrical preferences, resulting in distances of 2.5–3.2 Å between donor and acceptor atoms, donor–hydrogen–acceptor angles around 130–180°,[18] and a directionality dictated by the positions of lone acceptor electron pairs.[19] The majority of hydrogen-bond donors and acceptors constitutes of nitrogen- and oxygen-centric groups (e.g. amines, carbonyls or esters) though other polar moieties such as π systems may also be involved. Whenever formally-charged entities (deprotonated carboxylic acids, protonated amines, guanidyl groups) partake in the interaction, charge-assisted hydrogen bonds (one ionized partner) or salt bridges (formal charges on both donor and acceptor) are formed. Finally – and of special importance for surface-exposed, solvated binding sites –, partially-buried solvent molecules may also partake in forming these interactions, leading to a common phenomenon of water-mediated hydrogen bonds.[20] Given the abundance of these groups in biomolecules and a net

beneficial effect of 0.5–3.0 kcal/mol (after deducting the desolvation and entropy costs with the gain increasing up to 5.0 kcal/mol whenever ionized partners are involved), hydrogen bonds are of immense importance. This is all the more so, given their impact on the specificity in various protein-ligand complexes, exerted through the rigorous conditions of satisfying desolvated donor/acceptor groups.[21]

Van der Waals interactions

Named after the Dutch physicist Johannes Diderik van der Waals, these interactions, comprise a class of relationships formed between permanent or induced dipoles of non-polar atoms. Although only weakly attractive for optimal arrangements of a singular atom pairs (around 0.5 kcal/mol resulting primarily from dispersive forces rapidly diminishing with increasing distance),[22] they often dominate short-range relationships. This is due to imposing harsh energetic penalties for lack of spacial complementarity within a system – resulting from Pauli's exclusion principle –,[23] sufficient to altogether void the possibility of forming an interface between biomolecules, independently from other energy constituents.

The dualistic, attractive–repulsive nature of van der Waals interactions requires unique mathematical models to be described and quantified. One of the most commonly employed of these models is the Lennard-Jones potential (also referred to as the 6-12 potential), illustrating a generally accepted distance-energy relationship for a pair of atoms [Figure 2.2].[24]

Interactions involving π systems

Around 60% of all phenylalanine, tyrosine, and tryptophan side chains, are found involved in parallel or perpendicular (referred to as T-shaped or edge-to-face) aromatic stacking.[25] The source of this phenomenon lies in the unique distribution and properties of the π electrons, resulting in the formation of polarized areas that can favorably interact with one another as well as other groups. The abundance of aromatic rings within protein binding sites has also made "arming" prospective ligands with complementary groups, a common practice in the design and optimization of drug-like molecules. Through numerous studies, primarily based on high-level computational *ab initio* methods, it could be derived, that the energetical benefit of forming these interactions is similar to weak hydrogen bonds, with a free energy decrease between 1.0 and 2.5 kcal/mol, depending on the interaction geometry and properties of both partners.[26] The stacking between unsaturated ring systems is however not the only interaction involving π systems. A substantially

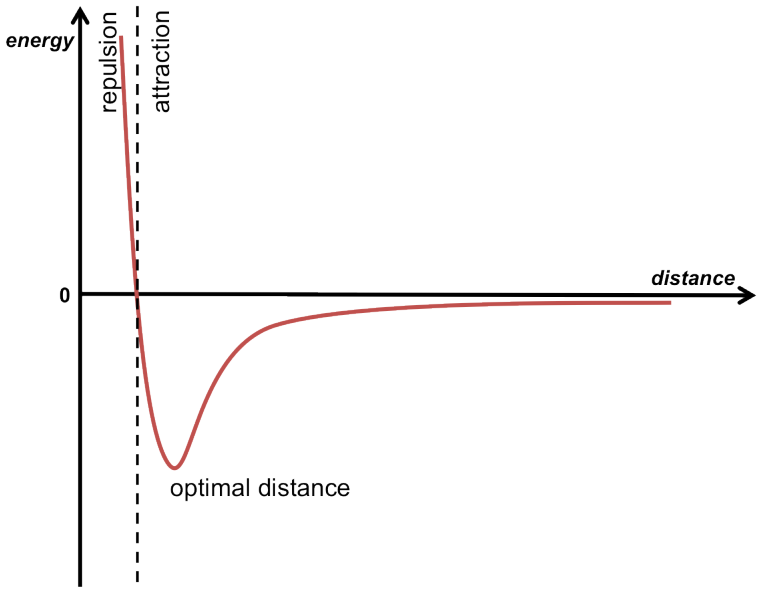


Figure 2.2: Schematic representation of the relationship between the van der Waals potential energy of two atoms and their respective distance. The highlighted optimal distance is typically 0.3-0.5 Å greater than the sum of the so-called atomic van der Waals radii defined from distances observed in crystal structures.

stronger – rivaling the energetical benefit of salt bridges – interaction may also be formed between π electrons of various moieties and groups bearing a formal positive charge. As with the π-π stackings, these so-called cation-π interactions have been successfully utilized in many design approaches, often targeting exposed side chains of arginine residues.[27] Attempts of integrating positively-charged groups into the structures of drugs remain less common however, due to the reported high promiscuity of such molecules.[28] It should also be mentioned that CH-π interactions may also benefit the binding affinities of various molecular entities, due to the attraction between the aliphatic groups and the π systems. While abundant in many complexes – including carbohydrate-protein pairings –, these interactions are however still challenging to quantify and optimize, with arguments supporting their dispersive, charge-transfer, or hydrogen bond-like nature constantly exchanged. Hence, a commonly utilized trait of π-electron related interactions, is that their strength increases along with the acidic/positive character of the partner group, leading to positioning various electron-withdrawing groups in their vicinity.

2.2.2 Entropy

Entropy is perhaps the single most elusive aspect of biomolecular binding. According to its statistical interpretation, it relates to the tendency of a given system to distribute the available energy between a number of conformational states. This definition dates back to the fundamental 19^{th} century work of Ludwig Boltzmann and relates to the probability of finding a given microscopical system in a distinct conformation.[29] Another, earlier definition, describes entropy as the overall energy related to all equilibrium and non-reversible motion within a system. This form – referred to as the thermodynamic interpretation of entropy – relies heavily on the works of Rudolf Clausius and often proves the more useful of the two with respect to understating the formation and stability of host-guest complexes.[30] In accordance with this theory, entropy is a state function determined by a set of parameters (e.g. temperature, volume, pressure), and not by the way a given state is acquired. The impact of temperature is perhaps the most profound, often determining entropy- or enthalpy-driven binding, or voiding the association altogether, with an "overheated" system "preferring" to stay decomposed.

Although, not uniformly defined – due to lack of means of its direct measurement –, the change in entropy is commonly described as consisting of three primary components

$$\Delta S_{tot.} = \Delta S_{solv.} + \Delta S_{conf.} + \Delta S_{r/t} \tag{2.5}$$

where $\Delta S_{solv.}$ corresponds to the solvent release and motion upon binding, $\Delta S_{conf.}$ relates to the loss of conformational flexibility with respect to the free solute molecules, and $\Delta S_{r/t}$ describes the loss of rigid body rotation and translation. Solvation entropy is commonly the largest constituent, given the significant interface surface areas stripped (desolvated) of water during all binding events. With the return of these molecules into the so-called bulk solvent commonly increasing their disorder, this component is also normally the only one favorably impacting the overall binding, providing as much as 40 kcal/mol of energy in 298 K.[31] The conformational entropy on the other hand is virtually always unfavorable, relating to the loss of rotational motion around the torsion angles of the protein and ligand. Still, the initial concepts of complete bond restriction upon binding – resulting in employing arbitrary "per-bond" penalties of 0.5–0.7 kcal/mol –, have ever since been proven inaccurate, with weak and surface-exposed interfaces retaining a significant degree of their free-state motion.[32] The entropy related to the rigid body movement, while significant, is commonly assigned arbitrary 2–5 kcal/mol constants (or ignored altogether) for the purposes of comparing various ligands of a receptor,[31] as long as their relative shape and size remain similar.

Taken together, the entropy constituents tend to follow the principle of "large numbers amounting to small changes", as often observed through Isothermal Titration Calorimetry (ITC) measurements. It also bears mentioning that a fourth constituent – the so-called vibrational entropy – is at times mentioned throughout literature. This quantity may however be merged with the other three, based on the theory that loss of translational and rotational motion results in the formation of additional vibrational degrees of freedom.[33]

2.2.3 Hydrophobic and solvation-related effects

Binding occurs in an aqueous environment. This obvious fact is at times easily overlooked, given too much focus on the already complex nature of protein-ligand interactions. In truth however, the thermodynamical events related to the water surrounding a host-guest complex – comprised of both enthalpic and entropic components –, are often determinant to the overall binding energy.

One reason for the importance of water – aside from the above–mentioned ability of forming bridged-hydrogen bonds – is its impact on the enthalpy of binding. Due to its small molecular size and high polarity, water molecules occupy virtually all receptor (and ligand) cavities, leading to the formation of the so-called first solvation shell. This also holds true for all binding sites, that require to be stripped of the surrounding solvent upon ligand approach. Depending on the character, shape, and size of these cavities – and as a consequence, the strength with which they interact with the solvent –, their desolvation may result in varying costs. Those are then further modulated, once the released water molecules rejoin the hydrogen-bond network of the bulk solvent. Due to such, the overall binding enthalpy can in fact be defined as the extent by which the protein-ligand interaction overcompensates the net cost of the solvent displacement.[34]

The water displacement itself is also closely related to the possibly most powerful factor determining all molecular rearrangement events (be it protein folding or a change of a single torsion angle), mainly the so-called "hydrophobic effect". Its origin – while still under discussion – is commonly related to the change in entropy accompanying the solvation/desolvation of various molecular surfaces. Specifically, upon surrounding a given solute, water molecules become more localized – the more so the more non-polar the solute –, leading to a substantial entropic penalty for loss of their inherent motion. This in turn pushes all solute molecules toward diminishing this cost through aggregating and minimizing the solvent-excluded volume. When looked at from the opposite direction, the hydrophobic effect may lead to substantial free energy benefits of returning waters to the bulk. Indeed, a common cor-

relation between the hydrophobic surface area buried upon ligand binding and the affinity toward a given receptor, has for a long time served as basis of many optimization approaches. Directly proportional to the heat capacity change within the system, this value is in fact one of the most crucial structural parameters employed in the design of various computational scoring functions.[35]

2.2.4 Entropy-enthalpy balance and compensation

Improving molecules through enabling more/stronger interactions is often met with little to no increase of the overall binding energy. This phenomenon arises from a very obvious fact that tighter binding inevitably leads to a decrease of the intrinsic motion within the system, and hence its overall entropy, resulting in a semi-linear relationship referred to as entropy-enthalpy compensation. While widely researched and accepted, this trend continues to pose difficulty to the field of medicinal chemistry, relating to the fact that various energy constituents (e.g. electrostatic interactions) rarely correlate with the binding affinity and cannot be optimized separately.[36]

Were it truly the case that entropy and enthalpy were perfectly synchronized, the overall binding energy would prove independent from the changes of either (caused for example by an increase of temperature), resulting in an impossible task of designing interactions without imposing restriction. While a trend is still observable in many cases, it does not hold true in general however, as proven by the numerous successful drug optimization attempts. Hence, a revised model of the enthalpy-entropy relationship is currently gaining attention [Figure 2.3], accounting for both the entropy importance in weak organic complexes (sub-micromolar affinities) and the enthalpy-driven association of "best in class" synthetic molecules.[37]

The accumulated knowledge regarding the entropy-enthalpy relationship, has also helped in establishing various efficiency metrics – with the most basic one being the result of dividing an estimated free enthalpy by the number of ligand atoms – that have proven valuable in biological and computational screening approaches.[38]

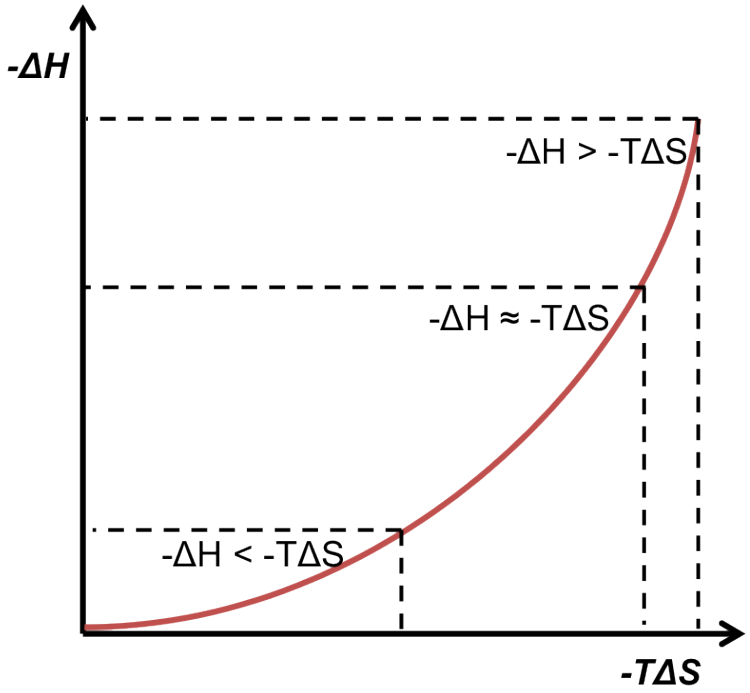


Figure 2.3: Idealized representation of the binding enthalpy-entropy relationship, illustrating a limit of the sustained entropy loss approached for very strong (possibly covalent) interactions. Figure adapted from[37].

2.3 Relevant computational methods

Experimental approaches probing aspects of host-guest association are often met with limitations regarding their cost and efficiency. Furthermore, a detailed energetic makeup of various systems can rarely be obtained outside of ITC experiments that consume precious amounts of substance and require interpretation in context of the specific measurement conditions.[39] To this end, the multitude of *in silico* tools developed and refined over the last few decades along with the gradual increase in machine processing power are of immense benefit. Below follows a short overview of computational methods geared toward predicting specific properties – hence the term "modeling" – relevant for structural drug design and optimization. A common practice in present-day research is employing multiple of these techniques alongside each other or coupling them with actual experimental procedures.

2.3.1 Structure generation

In silico methods of predicting protein folding and large-scale rearrangements (commonly centered around template-based homology modeling and extensive molecular dynamics simulations) have made tremendous progress in recent years.[40] Even so, obtaining a purely-computationally derived structure composed of dozens or hundreds of residues remains a formidable challenge. On the other hand X-ray crystallography still struggles with expressing certain proteins and resolving their structures with sufficient resolutions – preferably below 1.2 Å to provide an atomic level of detail.[41] Luckily, for the purposes of exploring ligand binding to a known receptor structure (e.g. resolved with a different ligand), the vast array of computational docking software often provides all that is necessary.

Computational docking consists of two basic components, sampling and scoring (the latter discussed in the next section). The first part corresponds to the generation of multiple putative ligand poses that fit a given receptor binding site. These poses are commonly generated by means of an exhaustive, automated conformational search, guided by either shape-matching or energetical criteria.[42] As did the entire field of structure-based design, the fully automated docking algorithms have throughout the years evolved away from the rigid "lock and key" concepts and now offer state-of-the-art flexible treatment of the ligand (*Glide*) or the entire binding interface (*Induced Fit Docking*, *Cheetah*).[43,44] The validity of these methods could be proven by means of various benchmarks, in which Root-Mean-Square Deviations (RMSD) below 1-2 Å from corresponding crystal structures could often be obtained.[45] If need be, the so called "manual" docking may also be employed, in which the user – taking advantage of his knowledge about the given system – adjusts the conformation of the small molecule or even the receptor. Finally, intermediate solutions like template-based (guided by an experimentally-derived ligand pose) or ensemble docking (generating binding poses to multiple configurations of the receptor) may also be employed.[46] Given this vast diversity and rapid processing speeds, docking methods are the starting point of most molecular modeling endeavors.

2.3.2 Energy calculation and ligand scoring

Obtaining a realistic representation of a protein-ligand configuration – while valuable – is often not sufficient to proceed with compound design or analysis. To this end, relative or absolute estimates of the binding energy (preferably decomposed into its various constituents) are required to interpret the structural data. For this purpose, an array of mathematical means of pre-

diction, referred to as scoring functions may be employed. Based on their origin and functional form, scoring functions are normally categorized between different groups, with force-field and empirical types being of primary interest in context of this work, along with the closely related Quantitative Structure-Activity Relationships (QSAR).[47] Given the varying case to case performance, approaches of employing multiple methods to yield a single, composite score are gaining increasing attention. These are referred to as consensus scoring.[48]

Force-field scoring functions

The underlying energetics of molecules may be obtained directly from the potential energy function present in most computational methods (including docking, in which they allow selection of the top poses within an ensemble). Along with sets of parameters derived by either means of high level *ab initio* calculations (themselves being inappropriate for scoring tasks due to tremendous computational requirements) or fitting to sets of experimental data, these functions are referred to as force fields.[49] Throughout the decades of intense improvement, force fields have become the basis of most molecular modeling, relied on by both academia as well as industry.

A common conceptual approach applied toward designing all potential energy functions may be presented using one of the most established among them – AMBER. In its basic, somewhat historical form,[50] the AMBER force-field (standing for "Assisted Model Building with Energy Refinement") consists of the following components

$$E_{\text{total}} = \sum_{\text{bonds}} k_r(r - r_{eq})^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{non-bond.} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}\right] + \sum_{\text{H}-\text{bonds}} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}\right]$$

with their meaning as follows:

First and second term represent the bond and valence angle energies of covalently-bonded atoms, using a harmonic approximation, with k a respective force constant parameter and r/θ – current and equilibrium bond/angle values.

Third term represents the torsional (ϕ) energies, with V_n a force constant, n the periodicity, and γ the equilibrium angle value.

Fourth term represents the non-bonded energy for *i*-*j* atom pairs separated by a distance *R*. Consists of van der Waals (Lenard Jones potential with *A* and *B* corresponding to specific atom types) and electrostatic (coulombic) energies.

Fifth term represents hydrogen-bonding, with *R* being the distance between donor and acceptor (*i* and *j*), and *C/D* – parameters corresponding to their atom types.

Although commonly enhanced with additional or improved terms (polarizability in *CHARMM*, angular hydrogen bond geometry and metal interaction terms in *Yeti*), force-field based methods display certain limitations.[51,52] Firstly, the distance-dependent dielectric treatment prevents accounting for various solvation effects. One way for remedying this limitation is an explicit (atomistic) representation of the solvent molecules. Still, the corresponding methods such as Free Energy Perturbation (FEP) or Thermodynamic Integration (TI),[53] commonly require coupling with molecular dynamics (MD) protocols along with extensive computation times. A compromise between accuracy and speed may be obtained through the employment of implicit solvation models. Among these, the Poisson-Boltzmann Surface Area (PBSA) and Generalized Born Surface Area (GBSA) models are especially popular and often integrated into standalone post-processing modules, such as MM/PBSA (software *Amber*) and MM/GBSA (software *Prime*).[54,55] Nevertheless, while proven to correlate well with various sets of experimental results,[56] the implicit solvation models do not remedy all the limitations of force-field based approaches. Specifically, the previously-mentioned correlation between various energy components (making them in principle non-additive), is often a source of severely overestimated *in silico* predictions. Also, the persisting lack of atomistic solvent properties may lead to inaccurate treatment (no hydrogen-bonding with the solvent) and prevents accounting for its entropic components.

Empirical scoring functions

While also relying on a summation of various energy terms, empirical scoring functions (e.g. *GlideScore*) boast certain advantages over their force-field counterparts.[57] Firstly – and as already implied by their name –, these functions consist of empirical energy components, weighted to fit so-called training sets of experimental data and allowing (in principle) higher predictive accuracy. Secondly, due to a very simple mathematical form

$$E_{\text{total}} = \sum_i W_i E_i \tag{2.6}$$

where E is a specific energy term and W its weighting coefficient, these functions prove highly efficient.[58] Lastly, given the not necessarily computational sources of parametrization, empirical scoring functions enable incorporation of solvation and entropy terms, difficult to account for outside of an experiment. Still, the effective performance of the corresponding methods, relies heavily on the quality and robustness of their underlying training sets. Due to retaining the energy decomposition scheme, double-counting (overestimation) errors may also persist.

Quantitative structure-activity relationships

Structure-Activity Relationships (SAR) comprise concepts driven by the idea that structural features of a system are related to its various properties.[59] Given the unique predisposition of computers in handling SAR-related data, design approaches are common to employ the so-called Quantitative Structure-Activity Relationships (QSAR) as a means of rapid and accurate affinity prediction.[60] Although not scoring functions *per se*, the connections between the QSAR methodologies and force-field/empirical concepts become apparent, with both common to employ the same basis of predictions (force fields, experimental data fitting, structural data mining). Nevertheless, it should be emphasized, that these methods are by far more specialized and diverse. Commonly QSAR methodologies are classified based on so-called dimensionality [Table 2.1], ranging from one-dimensional relationships between affinity and an experimental observable (pK_a, $\log P$), to multidimensional (mQSAR) concepts, employing three-dimensional structural ensembles alongside multiple receptor models (*VirtualToxLab* platform).[61]

Table 2.1: Dimensionality of QSAR techniques in relation to the prediction of a given property (commonly ligand affinity).

Dimensionality	Concept
1D	Relationship between an observable and the property
2D	Relationship between a structural pattern (pharmacophore) and the property
3D	Relationship between the three-dimensional structure and the property
4D	Incorporation of multiple representations of each molecule (conformations, protonation states, isomers)
5D	Incorporation of receptor-related events such as induced-fit (flexible and/or multiple binding site models)
6D	Incorporation of multiple solvation representations

As with the empirical scoring functions, QSAR models are commonly established based on biological and structural data of a training set. However, while retaining the underlying benefits of accuracy and accounting for elusive entropy and solvation properties, QSAR techniques often venture further with respect to their usefulness. Specifically, the higher dimensionality approaches (4D+), enable eliminating the human bias of selecting the receptor/ligand conformation from an ensemble of poses.[62] Furthermore, at the pinnacle of their sophistication, the models may even shed the molecular mechanics "ball and spring" representation altogether, resulting in flexible receptor surrogates ("envelopes") with various properties mapped on their surface.[63] On the downside, the technical complexity of mQSAR methods along with the strong relation to the underlying experimental data, often makes them laborious to employ and inappropriate toward handling molecules strongly deviating from their initial training sets.

2.3.3 Insight into structural dynamics

The value of molecular docking and scoring cannot be denied, having served numerous applications throughout the last three decades. Nevertheless, while commonly coupled with state-of-the-art sampling and energy minimization algorithms, these methods cannot alone account for all effects of the motion accompanying binding events. On the other hand, the experimentally-derived structures carry a strong relation to the conditions in which they are procured

(involving low temperature and presence of crystallization additives), leading to the occurrence of crystal packing effects. These techniques also preclude insight into the inherent motion of biomolecules (whenever it transcends the boundaries of vibrational movement or rotation around unrestricted bonds), given the employed averaging of time and space.[64] Due to such, molecular dynamics (MD) simulations are perhaps the single most important *in silico* contribution toward the field of structure generation, refinement, and analysis.

Conventional molecular dynamics

From a theoretical viewpoint, molecular dynamics rely on the iterative numerical solution of Newton's equation of movement

$$\vec{F} = m\vec{a} \tag{2.7}$$

where F is the force exerted on an atom, m its mass, and a the acceleration.[65] Based on a relation with the potential energy of the system, the force may also be expressed as follows

$$\vec{F} = -\frac{dE_P}{dr} \tag{2.8}$$

where r corresponds to displacement (e.g. change in the Cartesian coordinates). Using the two equations, a relationship between the change in position, the energy, and the time (t) may be established

$$-\frac{dE_P}{dr} = m\frac{d^2\vec{r}}{dt^2} \tag{2.9}$$

enabling calculating the configuration of a system at any point in time, based only on an initial set of coordinates (commonly obtained from crystal structures) and velocities (generated by the algorithm).[66]

In practice, it is common for MD simulations to retain the same molecular mechanics principles as the docking protocols, including deriving the underlying energetics by means of force fields. Alongside the exploration of various structural events, this enables energy quantification (albeit as an average over an ensemble of poses referred to as the trajectory) with most of the "static" methods presented earlier (such as MM/GBSA). If need be, more precise calculations (FEP, TI or even quantum-level simulations, known to achieve experimental-level precision) may also be coupled with the algorithms.[67] Given the tremendous increase of computational power, simulations using an explicit representation of the surrounding water molecules have also become standard, enabling insight into both their enthalpic and entropic

contributions. Finally, techniques such as normal mode analysis (NMA) or quasi-harmonic/harmonic approximations may be applied to the ensembles of MD frames, yielding estimates of translational, rotational and conformational entropy.[68,69]

Taken together, the multitude of MD techniques makes them a robust and multi-purpose means of structural prediction, refinement, and scoring, well worth the additional time devoted to their preparation and running. Still, the applicability of these techniques relies heavily on the time scale of the event we wish to observe [Table 2.2]. Due to such, while local reorientation of side chains accompanying ligand binding (commonly ns-scale) may be simulated using relatively inexpensive desktop machines, events such as conformational selection or allosteric transition (μs to ms), remain inaccessible without access to multi-million dollar supercomputers.[70]

Table 2.2: Approximate ranges and time scales of various structural events.[71]

Event	**Extent of atomic motion [nm]**	**Time scale [s]**
bond vibration	0.001–0.01	$< 10^{-12}$
rotation of exposed side chains	0.5–1.0	$\sim 10^{-9}$
rotation of buried side chains	~ 0.5	$\sim 10^{-6}$
allosteric transitions	0.5–4.0	$10^{-6} - 10^{-3}$
loop motions	1.0–5.0	$10^{-9} - 10^{-6}$
domain motion/association	$\gg 1.0$	$> 10^{-6}$

Advanced sampling techniques

An alternative and far more feasible foothold toward exploring rare/large-scale conformational events may be offered by several specialized (at the expense of detailedness and scoring capabilities) MD algorithms developed primarily within the last decade. These techniques may be assigned between two major classes:

- Methods prioritizing sampling of specific degrees of freedom (e.g. respective protein domain motion), while rigidifying or ignoring the others.
- Methods modifying the sampled potential energy landscape, allowing "escaping" the local minima in which the systems spend most of the time.

A common example of the first type of techniques is the *SHAKE* algorithm,[72] imposing constraints on either all or only hydrogen-related covalent bonds

within the system. Through this, the necessity of accounting for bond motion is eliminated, allowing the simulation to progress in larger time steps. A further case are the various "coarse-grained" methods, in which groups of atoms (residues, domains, etc) are treated as single rigid "beads", with the simulation sampling only their relative motion.[73] Finally, a class of intermediate methods employing strong forces driving the simulations, also exist (umbrella sampling, metadynamics, steered molecular dynamics), proving very beneficial toward cases when the endpoints of a given conformational change or its direction (definable by means of so-called collective variables) are already known.[74,75,76] Regarding the second type of techniques, they correspond to a family focused on generating conformations at the expense of their realistic distribution. A concept employed in simulated annealing and replica exchange techniques for example,[77] is increasing the simulation temperature resulting in higher residual motion. Alternatively, methods based on biasing potentials (applying "boost" levels inversely proportional to the potential energy) also exist, with accelerated molecular dynamics (AMD) being of primary importance for this work and described in the Materials and Methods section.[78]

2.4 The FimH adhesin

2.4.1 Specifics of carbohydrate-protein binding

Carbohydrates are among the most abundant biomolecules (alongside nucleic acids and proteins), constituting the external cell surfaces of every living organism. Although consisting of a relatively narrow set of base molecules (monosaccharides), they form a diverse array of complex structures, allowing them to conduct numerous structural, recognition, and metabolic functions. This broad functionality is enabled through the formation of variously-positioned glycosidic linkages between the monosaccharide building blocks, leading to the formation of extensive branched biomolecules. In addition, the base components themselves may exist in open, "chair", or "boat" conformations (with the latter two displaying α and β isomerism on their anomeric center), further extending the structural and functional diversity of carbohydrates.

In terms of interaction with other biomolecules, poly- or oligosaccharides (commonly referred to as glycans) tend to bind to highly-specific protein receptors, such as lectins and sulfated glycosaminoglycan binding proteins (SGAGs).[79] The formation of these complexes is also often involved in various pathophysiological events like bacterial and viral infections, inflammatory

cascades or cancer metastasis.[80] Due to such, and in no small part thanks to the insight provided by the various experimental and computational methods presented in the previous sections, various carbohydrate receptors have become a vital target of pharmaceutical research.

Although relatively simple in terms of chemical composition, small natural carbohydrates are not a suitable basis for drug design. This is largely due to their high polarity preventing passive permeation through lipid layers and, as a consequence, oral bioavailability. When administered through alternative means (e.g. injected), the compounds also suffer from rapid renal excretion. In order to address these drawbacks, a novel, largely untapped class of small molecules, known as glycomimetics is more commonly the center of attention. Still, the design and optimization of drug-like molecules targeting a given carbohydrate receptor remains all but simple. For instance, while the issue of high polarity is theoretically circumventable by the elimination of non-essential hydroxyl groups, many of the carbohydrate receptors display an extreme degree of specificity toward their natural binders, preventing even the slightest deviation form the original structure. Despite progress in the field, the synthesis of these substances is also challenging and lengthy. Hence till date, the tremendous effort on both academic and industrial fronts could only yield a limited number of market-approved carbohydrate-based drugs [Figure 2.4].[81]

From a viewpoint of computational design, carbohydrate-protein binding presents no less of a challenge [Figure 2.5]. Firstly, the high content of polar and non-polar groups alike, requires force fields to correctly account for an interplay of inter- and intra-molecular interactions. Of particular priority is the treatment of all hydroxyl groups that (both in bulk solvent and the carbohydrate binding pockets) participate in the formation of a geometrically rigorous network of hydrogen bonds. Equally important are the non-specific electrostatic interactions formed by these polar groups, that have to be modeled in context of long range and shielding (i.e. partially diminishing) effects of binding site residues. While commonly of lower magnitude, van der Waals/hydrophobic contacts enabled by the non-polar elements of the carbohydrate rings have also be taken into account. This includes CH-π interactions formed with – primarily – phenylalanine and tyrosine side chains, commonly exposed into the binding sites.[82] In addition, the distinct geometries of glycosidic linkages between the carbohydrate subunits are rarely covered by the base parameter sets, calling for the employment of specialized force fields such as GLYCAM.[83] Finally, and perhaps presenting the greatest challenge yet, the high inherent conformational motion of carbohydrate structures (hydroxyl groups and glycosidic linkages, “chair flips”, hydrogen-bond exchange), requires state-of-the-art conformational searching algorithms or

molecular dynamics simulations to be accounted for. This necessity is pronounced by the commonness of forming highly-flexible interfaces on protein surfaces, that cannot be appropriately described without accounting for entropic and solvation contributions.

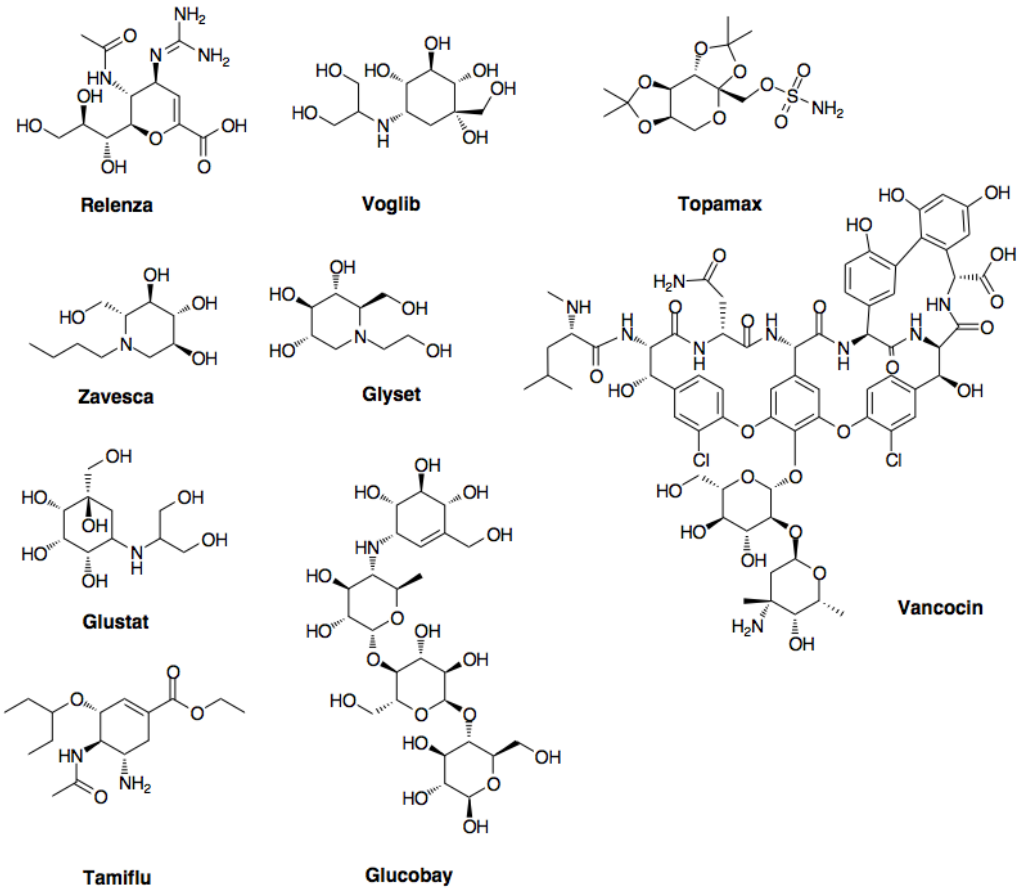


Figure 2.4: Examples of carbohydrate-based drugs along with their market brand names.

Taken together, while still uniquely predisposed toward handling the above-mentioned hurdles, modeling methods aimed at carbohydrate-protein binding require utmost attention and knowledge about a given system to prove of benefit. Due to such, a general carbohydrate scoring function (if at all obtainable) has yet to be introduced.[84]

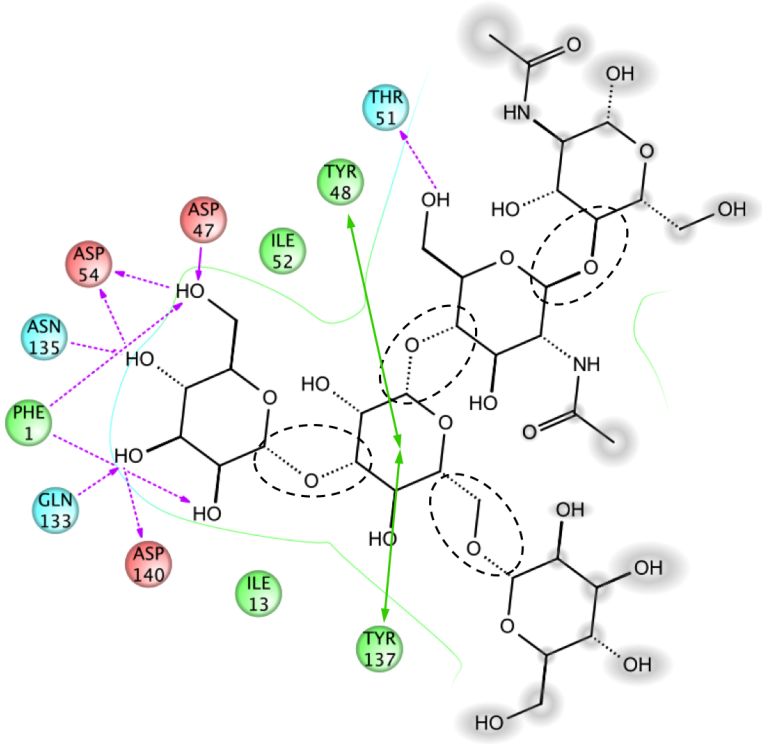


Figure 2.5: Scheme of a carbohydrate-protein interface, based on a structure of the FimH adhesin with its natural ligand (specifics discussed later in this work). Depicted are the following interactions: hydrogen bonds (violet), charge-assisted interactions (red), polar interactions (blue), non-polar interactions (green with arrows corresponding to CH-π contacts), solvent exposure (gray), and covalent glycosidic linkages (dashed ovals).

2.4.2 FimH: context and role

Urinary tract infections (UTI), caused primarily (70–95% of cases) by uropathogenic *Escherichia coli* (UPEC), are a common health issue affecting millions of people each year and accounting for extensive medical costs. Although normally non-severe (with developments of kidney or prostate diseases mostly related to untreated cases), symptomatic UTIs still result in a substantial decrease of life quality. Particularly affected are women, out of which approximately 40–50% experience at least one UTI in their lifetime, with a high chance of relapse.[85] While the commonly employed β-lactam (e.g. amoxicillin) and fluoroqiniolone (e.g. ciprofloxacin) antibiotic treatments often prove effective in short-term countering the UTI symptoms, they are more and more often met with the development of bacterial resistance upon pro-

longed/repeated application. As a consequence an efficient, non-antibiotic, and orally-available treatment strategy is of high demand.[86]

A potential foothold toward countering UTIs, relates to the initial stage of UPEC host invasion – the adhesion to the urinary bladder. This step is mediated through the prevalent fimbriae expressed on the bacterial surface, the multimeric type 1 pili [Figure 2.6]. A dominating structural feature of these adhesive organelle is the presence of up to several thousands of non-covalently bound, repeated FimA subunits, resulting in pilus rod lengths in the μm scale. A further characteristic is the presence of singular FimF, FimG, and FimH subunits on the tip of the pili. The assembly of the rod is handled in participation of the FimC chaperone protein, responsible for temporarily stabilizing the respective subunits through donation of a missing C-terminal β-strand. Upon assembly completion, this subunit is detached, with the donor strands provided instead by each preceding domain.[87]

Among the pili subunits, the terminal FimH protein has proven to be the primary UTI virulence factor, directly responsible for bladder cell adhesion and resistance to natural clearance mechanisms.[88] This makes it a promising target of inhibition by means of molecules mimicking its natural ligands – the α-mannosidic epitopes expressed on the host cell surfaces. Given the lack of selective pressure imposed on the invading bacteria, such inhibition (preventing the entire infection cycle) would pave the way to a novel UTI treatment method, unlikely to encounter resistance phenomena.

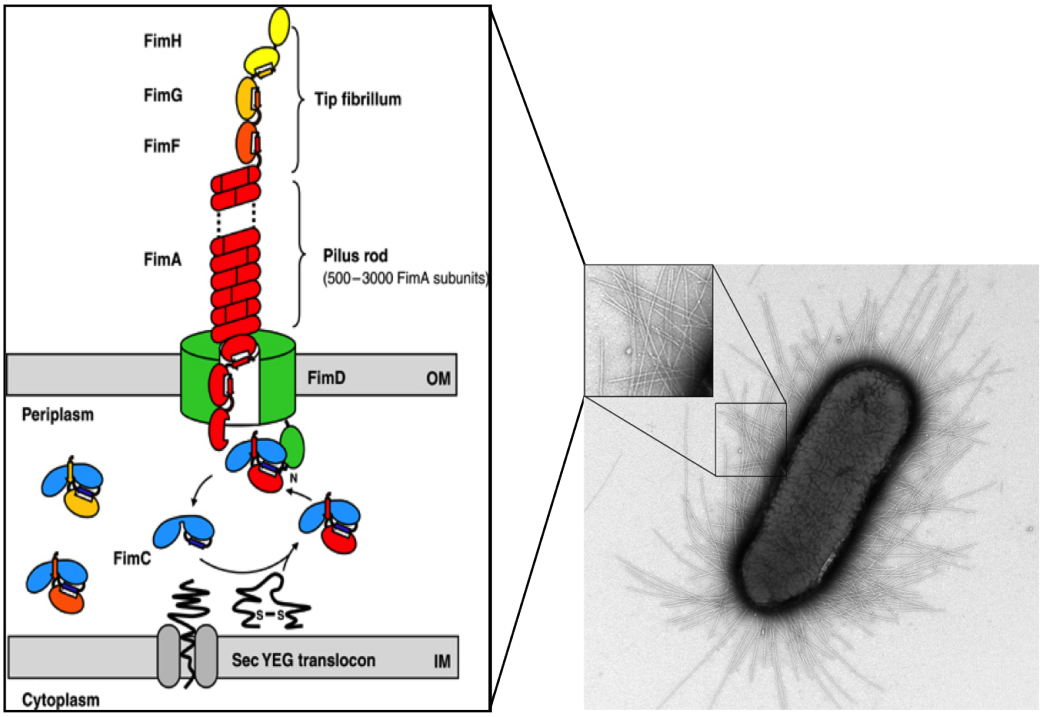


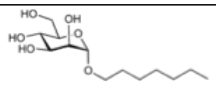
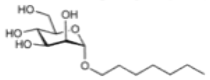
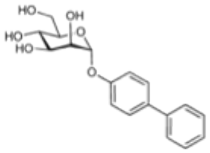
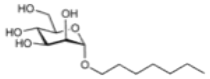
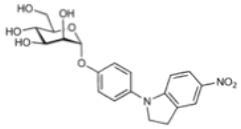
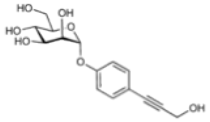
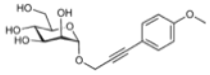
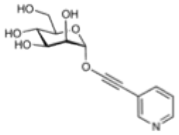
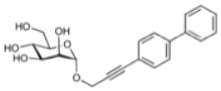
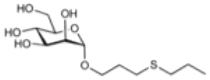
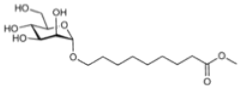
Figure 2.6: (left) Schematic representation of a type 1 pili and its assembly, on the *E.coli* surface (right). Images adapted from [89] and [90].

2.4.3 FimH structure

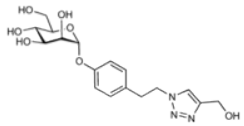
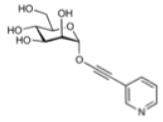
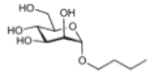
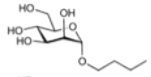
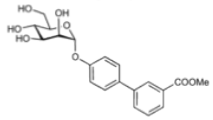
Nearly three decades ago, Sharon and co-workers explored various mannosidic antagonists for FimH-mediated bacterial adhesion, obtaining affinities in the milli- and micromolar ranges.[91] The first crystal structure of FimH was solved in 1999,[92] followed by a first holo (containing a bound D-mannose) structure in 2002.[93] Since then, an impressive amount of FimH structural data has become available in the Protein Data Bank (PDB) [Table 2.3]. Through these, it could be revealed that the 29 kDa FimH protein was constituted of 279 amino acid residues forming – unlike the remaining pili subunits – two domains. These were the N-terminal lectin domain (LD; residues 1-156 including the ligand binding site) and the C-terminal pilin domain (PD; residues 160-279), connected by a short linker strand (residues 157-159). In addition, a donor strand necessary to stabilize the PD, would commonly be provided by a FimC protein wedged between the FimH domains. Later on, structures of stable, detached LDs were also procured with various ligands.[94] A striking feature of all but one of the structures – independent from the presence of the PD and FimC –, are the nearly identical conformations of the LD, its binding site, and (if present) the mannose moiety of the bound

ligand. The sole structure deviating from this trend, contains the entire pili tip (including domains FimG and FimF in addition to FimH and a repositioned FimC), and provides an entirely new context for discussing FimH-ligand binding, tackled in Section 4.3.[95]

Table 2.3: Overview of solved FimH crystal structures. Initial five entries correspond to structures referenced in this work but not yet published nor deposited in the Protein Data Bank. Structure resolutions are given in Å. Unless stated otherwise, all ligands (if present) were co-crystallized with the proteins.

Code	Ligand	Resolution	Remarks
4BUQ		2.2	–
–		1.4	pending deposition/Y137A mutation
–		2.5	pending deposition
–		2.8	pending deposition/Y48A mutation
–		1.2	pending deposition
4AUU	–	1.6	apo
4AUY		2.1	–
4AV0		2.1	–
4AV4		1.9	–
4AV5		1.4	–
4AVH		2.1	soaked structure
4AVI		2.4	soaked structure

Continued from previous page ...

Code	Ligand	Resolution	Remarks
4AVJ		2.1	soaked structure
4AVK		2.4	soaked structure
3JWN	–	2.7	apo/two domains/low-affinity protein
2VCO	oligommannose-3	2.1	–
1TR7		2.1	–
1UWF		1.7	–
3MCY		2.9	–
1KLF		2.8	two domains/high-affinity protein
1QUN	–	2.8	apo/two domains

The dominant crystal conformation of FimH is convenient to present using the work of Wellens *et al.* from 2008. The corresponding structure contains an oligomannose-3 ligand, bearing a Manα1-3Manβ1-4GlcNAc motif, proven determinant for natural FimH adhesion [Figure 2.7].[96] As with the other structures, the isolated LD of this structure, consists primarily of an ensemble of anti-parallel β-sheets connected by loops on both domain ends. A solvent-accessible, surface binding site is formed at an interface of three of these loops (identifiable by the presence of residues Ile13, Tyr48, and Tyr137 respectively). A closer inspection of the site reveals a terminal mannose moiety of the oligomannose-3, tightly anchored in the pocket by an extensive network of hydrogen bonds. The remaining sections of the motif adapt a relatively planar conformation, inserting between the two parallel-aligned tyrosine residues lining the pocket's periphery. This structural element of FimH is commonly referred as the *tyrosine gate* or – upon inclusion of additional neighboring residues – the *hydrophobic rim*. Analyses of further structures complexed with monovalent α-D-mannoside mimics, present a very similar

picture, with various aglycones attached to the carbohydrate taking over the tyrosine-stacking role of the central oligomannose-3 moiety. The parallel alignment of the tyrosines is also maintained in all but one apo structures (aside from the effects of crystal packing). This unique complex, obtained by Han and co-workers in 2010, presents a rigid, biphenyl α-D-mannoside enforcing a perpendicular Tyr48-Tyr137 alignment.[97] It is thus assumed that the Tyr48 residue is prone toward taking on different conformations depending on the specifics of a ligand.

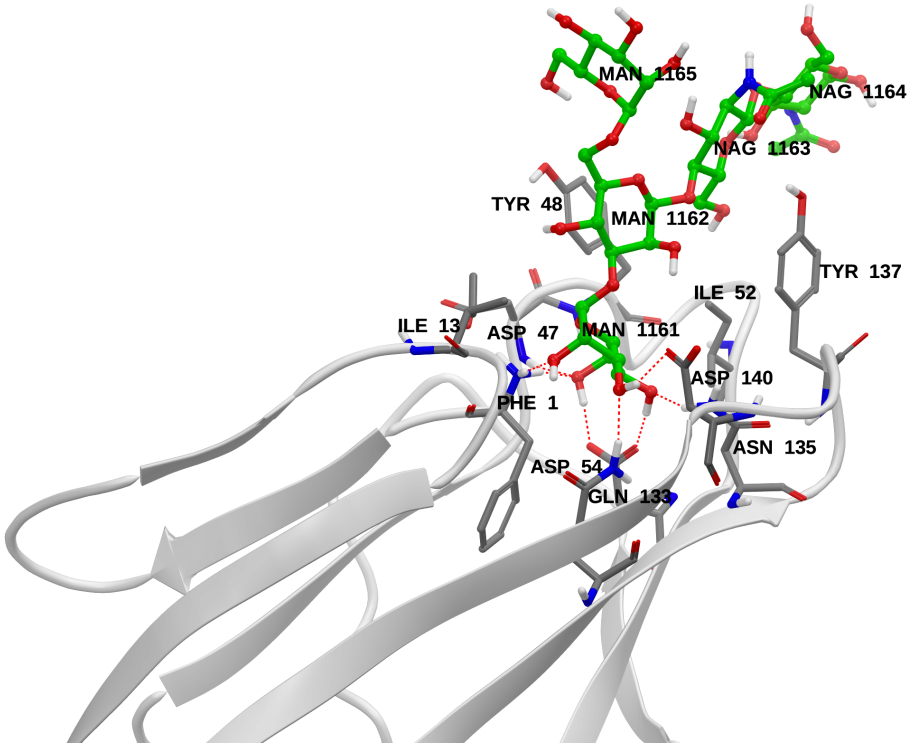


Figure 2.7: FimH lectin domain complexed with part of the natural high-mannose epitope – oligomannose-3 (PDB structure 2VCO with 2.1 Å resolution and added hydrogen atoms). A terminal mannose moiety is rigidly anchored by a tight network of hydrogen bonds, whereas the remaining units form a planar conformation, partially inserted between two tyrosine gate residues.

Based on the abundance of structural and biological data, two primary strategies could be established toward further improvement of mannosidic ligand affinity. The first one corresponds to employment of multivalent α-mannosidic structures, resembling the natural high-mannose epitopes and aimed at taking advantage of either secondary binding sites (not identified till date) or multiple pili expressed in close proximity.[98] The second approach

relies on rational design based the paradigm of further optimizing the contacts between the aglycones of monovalent α-D-mannosides and the *tyrosine gate* (necessary given the low affinity and selectivity of the mannose alone). This concept relates to one of the central points of this thesis and is discussed starting from Section 4.1.

2.4.4 Further FimH properties

Before proceeding with the discussion concerning mannosidic ligand optimization, certain unique properties pertaining to the structure of the entire FimH protein need to be highlighted.

Firstly, owing the non-rigid connection of its subunits resulting in the existence of at least two distinct conformations, FimH may exhibit either slip- or catch-bonding characteristics. This property, has been confirmed by flow chamber assay experiments,[99] and attributed to increased/decreased protein-ligand complex lifetimes under various conditions. The first, native conformation, is assumed under lack of urine flow and stabilized by a contact between the PD and LD. The second, visibly extended conformation exists primarily under flow conditions, with the LD firmly attached to the mannosylated cell surface. In this case, the shear force imposed on the bacterium causes the two FimH domains to separate (aside from the short strand of residues 157-159). As a consequence, several LD loops natively involved in interactions with the PD are liberated, leading to an untwisting of a central β-sandwich fold and elongation of the entire domain.[95] With this effect also allosterically transmitted to the mannose binding site on the tip of the LD – causing it to significantly tighten upon shear application –, the two conformations of FimH are commonly referred to as low- and high-affinity states. The reassumption of the compressed structure upon loss of shear or dissociation from the cell surface, is commonly known as self-inhibition. In general, the catch-bonding property – observed also for other systems such as selectins – is considered to be a natural evolutionary mechanism, enabling UPEC mobility within a host environment while at the same time preventing clearance by natural processes (i.e. urine flow).[100]

A second characteristic of FimH, is an array of natural mutations, involving at least 15% of all residues.[101] This feature, related to maintaining maximum pathogenicity in various conditions, results in the existence of *E. coli* strains expressing FimH variants with differing mannose-binding capabilities. While structural data on the various variants are scarce, it is commonly argued, that their properties relate to the ease with which they may transit between respective affinity states, i.e. a balance of a pre-existing conformational equilibrium. In extreme cases (such as mutation of residue 62) the

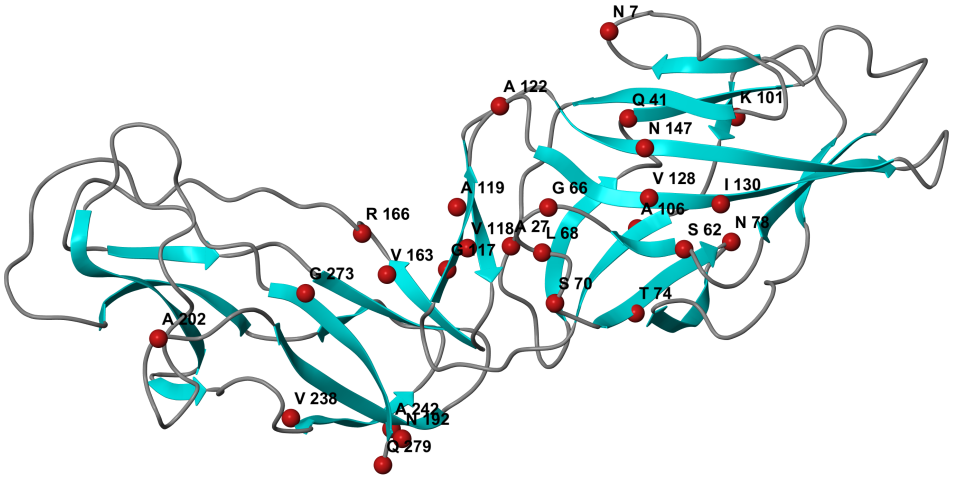


Figure 2.8: Two domain structure of the F18 FimH variant (native compressed state; PDB structure code 3JWN, 2.7 Å resolution). Residues commonly susceptible to point mutations marked with red beads.

shift of this equilibrium may lead to a permanently "locked" high-affinity state. This theory is well supported by the concentration of most mutations around the PD-LD interface (where they can disrupt the domain interaction), with none found within the conserved binding pocket [Figure 2.8]. Although strongly related to prospective inhibition attempts, the only so far documented method of efficiently assessing the FimH domain conformation, is the binding of a selective LIBS antibody to an epitope in the domain interface (residues 29 and 152 to 157), accessible upon domain separation. Due to such, alternative means of determining the FimH state and properties for a given bacterial strain – especially related to no-flow conditions expected for inhibitor binding –, are of utmost necessity.

3. Materials and methods

3.1 Structure preparation

All of the presented work was based on either one or several of the following FimH crystal structures retrieved from the Protein Data Bank (PDB):

- 1UWF structure, 1.7 Å resolution. All protein residues retained. Used for the majority of all docking.
- 3MCY structure, 2.9 Å resolution. All residues of chain A retained.
- 3JWN structure, 2.7 Å resolution. Chain H along with 13 residues of chain G (FimG donor strand) retained.
- 1KLF structure, 2.8 Å resolution. Chain B along with 13 residues of the FimC chaperone (chain A) retained.
- 2VCO structure, 2.1 Å resolution. Chain A only.

Unless stated otherwise, only the above-mentioned protein residues, the co-crystallized ligands, and a single structural water molecule were used. All systems were pre-processed using an identical protocol that included addition of missing side chains (if necessary), calculation of hydrogen atom positions, hydrogen-bond network optimization, and energy minimization (with backbone constraints) with the *Protein Preparation Wizard*.[102] Protonation states were assigned using *PROPKA* for a slightly acidic pH of 6.5, reflecting urine.[103] Whenever modification of a specific side chain was necessary, the new orientations were assigned based on the most common rotamer found in the Schrödinger software library,[104] that did not clash with any other atom of the solute.

In addition, structures co-crystallized with heptyl (wild type, Y48A, and Y137A FimH), indolyl, and squaric acid derivatives are mentioned. These have not yet been published or deposited in the PDB and were only used as reference. For details regarding them please refer to Dr. Julie Bouckaert and Dr. Roland Preston.

3.2 Docking

Outside of the *VirtualDesignLab* (discussed below), all ligand docking was performed using the *Induced Fit Docking* (IFD) protocol,[105] employing flexible treatment of the ligand and all residues within its 8 Å vicinity. The initial compound geometries were constructed using the *Maestro* interface (*Maestro*, version 9.3, Schrödinger, LLC, New York, NY, 2012) and optimized (minimization and protonation state assignment for a pH of 6.5) using *LigPrep* (*LigPrep*, version 2.5, Schrödinger, LLC, New York, NY, 2012) and *Epik*.[106] The IFD protocol consisted of an initial phase of docking to the rigid binding site (designated using residues Phe1, Tyr48, Asp54, and Tyr137), followed by protein side-chain optimization, and re-docking to the optimized grids. Ligand scores were obtained from the underlying *Glide SP* function employing *GScore*[107]

$$GScore = 0.05\ vdW + 0.15\ Coul. + Lipo + Hbond + Metal + Rewards + RotB + Site$$

where *vdW* and *Coul.* are the van der Waals and electrostatic energy components, *Lipo* the contribution of the non-polar solvation, and *Hbond* – the hydrogen-bond energy. The remaining terms correspond to (in order): solute-metal interactions (if present), rewards and penalties for features such as buried polar groups or hydrophobic enclosure, penalties for freezing rotatable bonds, and non-directional polar interactions in the binding site. In addition the IFD protocol applied a score correction based on the extent of receptor reorientation necessary to obtain a given pose (i.e. the induced-fit).

In each case, up to eight docking poses (if found within a 15 kcal/mol window from the top-scored one) were retained for interactive inspection. From these, the result with the highest score that also reproduced the hydrogen-bond network around the mannose moiety was used.

3.3 Molecular dynamics

Depending on the system (all originating from crystal structures or docking), molecular dynamics simulations were carried out either with *Desmond* (using the OPLS 2005 force-field as implemented in the Schrödinger 2012 suite) or *Amber12* (employing the AMBER 99SB force-field).[108,109] In either case, the systems were solvated using an orthorhombic, TIP3P water box with periodic boundary conditions at a minimum distance of 10 Å from the solute.[110] Na^+ and Cl^- ions were added to neutralize the charges and account for physiological salt concentration (0.15 M). The systems were then equilibrated

using the default *Desmond* relaxation routine or – for *Amber* – a self-prepared protocol consisting of constrained and unconstrained energy minimization followed by a gradual temperature increase from 100 to 300 K over a period of 25 picoseconds (ps). Long-range electrostatic interactions (cut at 8.0 Å) were handled using the particle mesh Ewald summation.[111] The SHAKE algorithm was applied to all hydrogen atoms.[72] Production runs were carried out using the Martyna-Tobias-Klein isothermal-isobaric ensemble (NPT) and the Nose-Hoover (Langevin for *Amber*) thermostat, maintaining a constant temperature of 300 K.[112,113] Unless stated otherwise, these runs covered the span of either 2 or 30 nanoseconds (ns; referred to as "short" and "long" respectively). Energetic and structural data were recorded in 4.8 ps intervals for a simulation time step of 2.0 femtoseconds (fs).

3.4 Steered molecular dynamics

The steered molecular dynamics (SMD) simulations apply an external physical force on a system, driving its conformational changes in a designated direction.[76] For the purposes of this work, the SMD protocol implemented in the *Amber12* software suite was used to increase the distance between the end of the FimH PD donor strand and the anomeric oxygen of the bound ligand from 81 to 116 Å. All simulations were based on the 3JWN structure with a manually-docked methyl α-D-mannoside. The donor strand stabilizing the PD was obtained by retaining the first 13 residues of the FimG domain also resolved as part of the structure. The simulations were performed over the span of 10 ns, at a constant velocity of 3.5 Å/ns, offering a compromise between CPU time and sampling detail. The effective driving force (F) at any given time point – exerted by pulling a dummy atom attached to the anomeric oxygen via a virtual spring – was determined according to the relation

$$F = -k\Delta x$$

where k is a defined force constant and Δx the spring extension. The selection of the former parameter was crucial towards the validity of the simulations and guided by the so-called stiff-spring approximation.[114] In light of it, the spring should be rigid enough to maintain a linear change of the reaction coordinate (distance between designated atoms) but still sensitive to the responses of the system. Based on multiple trial runs with various settings, a force constant of 10.0 kcal/molÅ^2 was selected. Aside from these parameters, all simulation settings were identical to the *Amber* MDs described above.

In order to probe the free energy changes accompanying the pulling (referred to as the potential of mean force or PMF), a total of 10 SMD simu-

lations were performed. Their starting points were extracted from a short, constrained MD with the initial docking pose. The underlying energy changes (ΔG) could then be derived by employing the Jarzynski relationship[115]

$$exp(-\Delta G/k_bT) = \langle exp(-W/k_bT) \rangle$$

where k_b is the Boltzmann constant, T the temperature, W the work (equal to $F\Delta x$), and the angle brackets – an indication of a mean over multiple simulations. The resulting PMF was plotted in a global (reaction coordinate vs. accumulated work) and a local form (using the work accumulated within arbitrarily-defined 1 Å intervals).[116]

3.5 Accelerated molecular dynamics

Accelerated molecular dynamics (AMD) are a representative of the so-called enhanced sampling MD techniques. Their aim is the simulation of rare structural events (loop/domain motion, ligand unbinding), inaccessible to conventional methods.[117] This functionality is obtained by modifying the potential energy ($V(r)$) surface sampled during a simulation, by raising the meta-stable minima (consuming most of the sampling time) while retaining its general shape

$$V(r)^* = V(r) + \Delta V(r)$$

where $V(r)^*$ is the recalculated energy and $\Delta V(r)$ – the so-called bias potential determining the extent by which the original surface is raised. As implemented in the *Amber12* software suite, this quantity is defined as

$$\Delta V(r) = \frac{(E_p - V(r))}{(\alpha_P + E_P - V(r))} + \frac{(E_D - V_D(r))}{(\alpha_D + E_D - V_D(r))}$$

with $V_D(r)$ being the dihedral component of the potential energy at a given simulation step. The parameters E and α are unique to the AMD simulations and obtainable from short conventional MDs. The first variable determines the threshold of total/dihedral energy above which no boost is applied and should be set up with caution not to destabilize the system. The second parameter defines (inversely) the strength with which the acceleration is applied. The presented form of the bias potential is referred to as "dual-boosted" and enables accelerating the movement in the entire system (facilitating e.g. the slow diffusive motions due to including the solvent molecules) while additionally "encouraging" the solute to explore different conformations (related to dihedral angle changes).[118] Given the non-specific

nature of the added potential, no prior knowledge regarding the conformational behavior of the system is required prior to employing the AMD protocol.

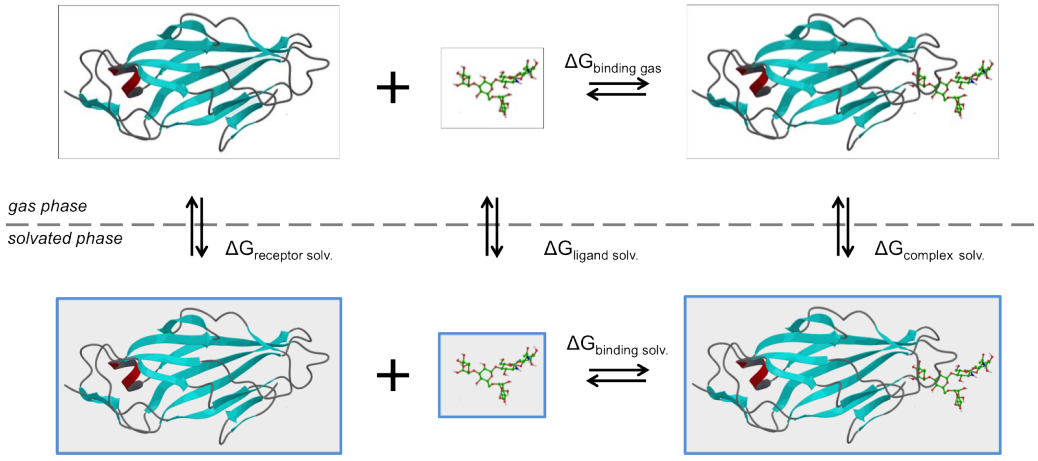
The simulated systems, were obtained based on the 1KLF and 3JWN FimH crystal structures, representing its known conformational states. As with all other two-domain simulations described in this work, these systems retained short sections of the accompanying FimC/FimG domains to stabilize the PDs. The acceleration parameters were derived based on multiple trial runs with different settings. Given the extent of conformational motion implied by the experimental structures, relatively large "boost" levels were selected, presumed to rank among the highest reported till date [Table 3.1].[119]

Table 3.1: AMD parameters used in the presented work. All values in kcal/mol and determined based on the energetics of short conventional MD simulations performed for each system.

System	$\boldsymbol{\alpha_D}$	$\mathbf{E_D}$	$\boldsymbol{\alpha_P}$	$\mathbf{E_P}$
3JWN (apo)	234	4107	8351	-122352
3JWN (holo)	234	4130	7698	-111365
1KLF (apo)	229	4051	7664	-112107
1KLF (holo)	230	4082	7681	-112221

3.6 MM/GBSA post-processing

The Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) method of free energy estimation presents a compromise between the processing speed native to docking and the accuracy of FEP/TI techniques, and has been extensively used in various design and screening approaches.[120,121] It relies on comparing the energetics of free and bound solute molecules as illustrated on the following scheme for the FimH adhesin



where the total binding free energy equals

$$\Delta G_{binding\ solv.} = \Delta G_{complex\ solv.} - (\Delta G_{receptor\ solv.} + \Delta G_{ligand\ solv.})$$

and (for receptor, ligand, and complex alike)

$$\Delta G_{solv.} = \Delta G_{gas} + \Delta G_{solv.} - T\Delta S$$

The so-called gas phase energetics (corresponding to the intra- and intermolecular interactions in the solute) are obtained from an underlying force field. The computation of the solvation energy changes relies on an implicit model and comprises polar (using the generalized Born model) and non-polar components (proportional to the solvent-accessible surface area changes accompanying the binding and calculated by the LCPO algorithm).[122,123] Aside from the GB model, other solvent representations are often employed, such as the Poisson-Boltzmann (underlying the MM/PBSA method) or Reference Integration Site model (RISM).[124,125] Finally, the (optional) entropy term is commonly provided by normal mode analyses (NMA) or harmonic/quasi-harmonic approximations.[126]

Given that each calculation requires only a single protein-ligand complex (with its components assumed to be valid representations of the unbound molecule conformations), the MM/GBSA protocols may be used to process structures originating from various experimental and *in silico* sources. In this work, either the *Prime* software implementation (using the OPLS 2005 force-field) or the *MMPBSA.py* module of the *Amber12* suite (using the AMBER 99SB force-field) were used.[127,54] With regard to the former, it was mainly employed toward post-processing of docking poses, improving the accuracy of their scoring. Residues within 8.0 Å from the ligand were treated as

flexible, providing (through a subsequent energy minimization) additional terms for protein and ligand strains. This method also employed the variable dielectric VSGB2.0 model to account for polarization effects.[128]. The *Amber* MM/GBSA employed the GB model developed by Onufriev *et al*,[129] and was used to obtain mean values over ensembles of MD snapshots (between 50 and 250 depending on the simulation length and purpose), diminishing the bias of single representations. These calculations also enabled decomposition of the resulting energy into per-residue contributions. For either software, explicit solvent molecules and ions were stripped to facilitate the computation.

3.7 Geometry analysis and clustering

The geometrical analyses of singular docking structures and *Desmond* MD simulations discussed in this thesis (distance and angle measurements, stability of hydrogen bonds) were performed using the in-build tools of the *Maestro* interface. For *Amber* simulations (MDs, SMDs, and AMDs), the *ptraj* module was used.[130] Hydrogen bonds were recognized for donor-acceptor distances below 3.0 Å and donor-hydrogen-acceptor angles above 135 degrees. Whenever necessary, trajectory frame clustering was performed based on the positions of protein C_α atoms or all heavy atoms of the ligand and the surrounding binding site (residues Phe1, Ile13, Asp47, Ile52, Asp54, Gln133, Asn135, and Asp140). The clustered snapshots were extracted every 10^{th} MD frame and grouped – using a hierarchical average-linkage algorithm – based on an RMSD cutoff of 2.0 Å between the clusters.[131] Volume measurements were performed using the *phase_volCalc* routine of the *Phase* software,[132] employing a spherical probe with a radius of 1.4 Å.

3.8 The *VirtualDesignLab*

The *VirtualDesignLab* is an *in silico* tool developed at our institute (based on the *VirtualToxLab* framework,[133] kindly shared by the Biographics Laboratory 3R) simulating and quantifying the binding of small molecules to a macromolecular target. Though initially developed for automating modeling tasks related to the FimH protein, it now allows re-orientation toward other targets. Its technology employs automated, flexible docking combined with multidimensional Quantitative Structure-Activity Relationships (mQSAR). Controlled over an easy-to-use graphical user interface and model-building tools, the *VirtualDesignLab* allows medicinal chemists for quick and straightforward design, screening and structural inspection of any compound of in-

terest.[134]

In order to provide a reliable *in silico* affinity estimate for a given system, it is necessary to account for protein-ligand interactions, solvation and entropic phenomena. With respect to the FimH adhesin, we utilized a set of 108 compounds along with their experimental affinity data, to develop and validate a corresponding mQSAR model (Appendix Table 7.1). When generating the model, the initial compound structures were constructed using the integrated model-building tool and then optimized with *MacroModel* (*MacroModel*, version 9.9, Schrödinger, LLC, New York, NY, 2012). Atomic partial charges were computed using the *AMSOL* package.[135] All structures were subjected to the conformational-searching algorithm *ConfGen*,[136] resulting in sets of low-energy conformations for each molecule in aqueous solution. Energetically feasible binding poses (within 10 kcal/mol from the lowest-energy structure) were identified by means of automated docking to two three-dimensional structures (*in* and *out* modes, cf. below) of the FimH carbohydrate-binding domain. The employed alignment (*Alignator*) and docking (*Cheetah*) protocols allowed for flexibility of both ligand and the protein (induced-fit) as well as dynamic solvation.[137,44] Several templates were used for the prealignment in order to account for distinct modes of binding reported previously. The underlying protein grids were obtained from the 1UWF and 3MCY crystal structures and pre-processed as described above. A total of 282 docking poses (allowing for multiple poses per ligand) comprising a 4D data set were then used as input (84 training and 24 test substances) for the mQSAR software *Quasar* to generate a series of quasi-atomistic binding-site models.[63] The underlying model families (comprising 200 members) were evaluated in consensus-scoring mode – along with a direct force-field scoring in *Cheetah* and the comparison of a molecule's interaction energy in a box of pre-equilibrated water and in the binding site. For validation, we additionally employed an alternative receptor-modeling concept *Raptor*,[138] featuring a substantially different scoring function (dual-shell induced-fit representation and introduction of a hydrophobicity term).

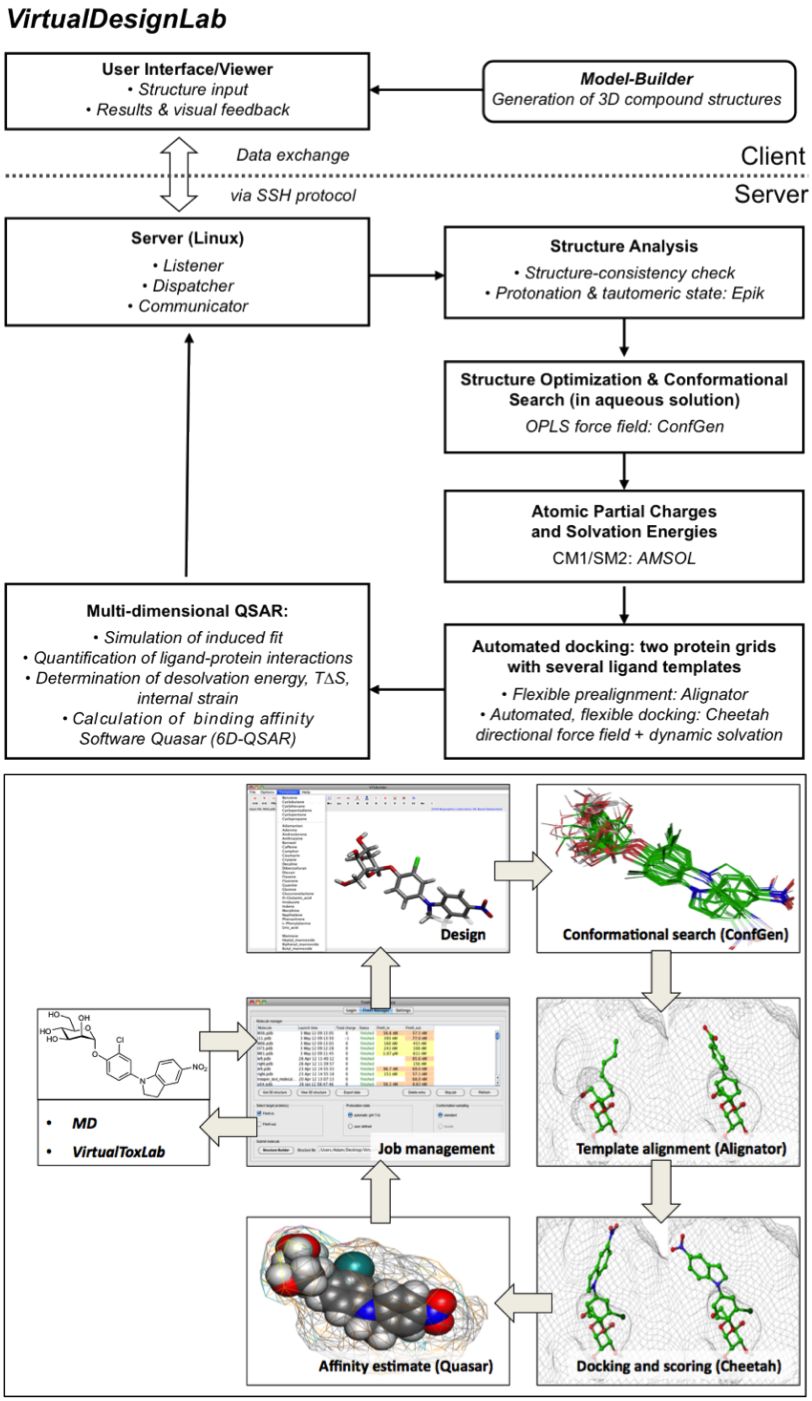


Figure 3.1: Example of a *VirtualDesignLab* workflow (bottom) and its flowchart (top). A compound of interest is designed within the framework or imported from an external source. Its initial conformation then undergoes an exhaustive conformational search, yielding feasible poses for subsequent docking and mQSAR. The entire workflow may be followed and controlled from a central interface window. Figures adapted from[139].

Every compound submitted to the *VirtualDesignLab* server (by means of imported PDB files or the integrated model builder) is subjected to identical protocols as those employed to train and validate the underlying mQSAR model(s) [Figure 3.1]. The affinity is calculated based on multiple components of the binding energy [Figure 3.2]. Protein-ligand interaction and internal strain energies (*Cheetah* and *Quasar*) are obtained using a force field with terms for polarization as well as hydrogen bond linearity and directionality.[52] The desolvation costs are calculated for the global minimum obtained from the conformational search, using a continuum solvation model. Loss of entropy is approximated from the number of rotatable bonds constrained upon binding to the protein. Induced-fit energy calculation is an inherent function of the *Quasar* algorithm. The affinity predictions are based on (up to) eight docking poses as obtained from *Alignator*/*Cheetah* (4D) and evaluate (up to) six induced-fit mechanisms (5D) and two solvation (6D) scenarios in order to account for the unique properties of certain binding sites (e.g. the surface-exposed FimH pocket). Protein-ligand structures may be viewed and/or downloaded (in PDB format) upon job completion. The files may also serve as input for other software including the *MD Client* also developed at our institute.

Figure 3.2: (top) Equation for calculating the binding energy used in the *VirtualDesignLab* and (bottom) the directional force field employed in *Cheetah* and *Quasar*. The individual terms (quantifying the bond lengths, bond angles, torsion angles, van der Waals contacts, geometries of hydrogen bonds, electrostatic/metal-ligand interactions, and ligand-protein polarization) are described in detail in a related publication.[52]

$$E_{binding} = E_{lig-prot} - E_{ligand\ desolv.} - T\Delta S - E_{ligand\ strain} - E_{induced-fit}$$

$$\begin{aligned} E_{\mathrm{lig-rec}} = & \sum_{\mathrm{bonds}} k_r(r - r_{eq})^2 + \sum_{\mathrm{angles}} k_\theta(\theta - \theta_{eq})^2 + \sum_{\mathrm{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] \\ & + \sum_{non-bond.} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{4\pi\epsilon R_{ij}}\right] + \sum_{metal\ pairs} \left[\frac{q_i^{CT} q_j^{CT}}{4\pi\epsilon R_{ij}}\right] \\ & + \sum_{\mathrm{H-bonds}} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}\right] \cos^2(\theta_{Don-H-Acc}) \cos^n(\omega_{H-Acc-LP}) \\ & + \sum_{metal\ pairs} \left[\frac{E_{ij}}{R_{ij}^{12}} - \frac{F_{ij}}{R_{ij}^{6}}\right] + \sum_{atoms} -\frac{1}{2}\alpha_i[E_i^0 E_i] \end{aligned}$$

3.9 Entropy treatment

Given the lack of suitable methods of rapid binding entropy prediction, a custom-made program was written as part of the presented work. Its aim was to correct a constant energetic penalty applied per each rotatable ligand bond, by probing the extent of motion still accessible to it upon complex formation. The conceived technology accepts the input of standard and extended (in-house developed format additionally storing atomic partial charges) PDB files, containing a protein-ligand structure. The underlying parameters were either derived from the *Cheetah* software (atom types, van der Waals radii, potential well depths), or fitted based on processing a diverse set of 317 protein-carbohydrate complexes discussed in the PhD dissertation of Sameh Eid.[140] In addition, various sources of input files were tested (*Induced Fit Docking* and *VirtualDesignLab* poses, extracted MD frames), to account for inconsistencies within the PDB file standard.

Upon recognizing the ligand and protein, the software maps their atom types, connectivity, bond orders as well as various elements that are not to be violated by later processing (4, 5, and 6-membered rings, inter- and intra-molecular hydrogen bonds and amide linkages). From there, non-rigid ligand torsions are identified and rotated (in 1 degree intervals) around their starting positions, using the Euler-Rodrigues formula [Figure 3.3].[141] The rotation is handled in four stages (180 degrees clockwise and counterclockwise on both sides of the bond), ensuring sampling of the entire conformational space. During each stage, the van der Waals energies of all atoms are checked (using the 6–12 potential model) to avoid steric clashes. Upon atom overlap, the program determines whether a rigid protein backbone, a proline or a cysteine residue (only when forming a disulfide bond) are involved, and either terminates the motion immediately ("hard" clash) or attempts to continue for a few additional degrees ("soft" clash). The distinguishing between the contact types enables accounting for protein side chain flexibility, often maintained upon binding.[142] An identical scan is then repeated for the ligand alone (starting from the bound conformation or a separately provided "global minimum"), to analyze intra-molecular restrictions. Overly-compressed poses are filtered out from the pool of found "virtual conformers", given the tendency of small molecular ligands to bind in extended geometries.[143] The number of remaining conformers is then used to scale down the cost of 0.7 kcal/mol per a fully frozen torsion at 298 K.[144]

$$T\Delta S_{Conf.} = 0.7 * No.\ of\ rotatable\ bonds * \frac{(free\ conf. - bound\ conf.)}{free\ conf.}$$

The resulting conformational penalty for all bonds is added to an estimate of entropy loss for rigid body translation, rotation, and vibration, obtained from an external RRHO (stands for Rigid-Rotor Harmonic-Oscillator) module of the *MacroModel* software.

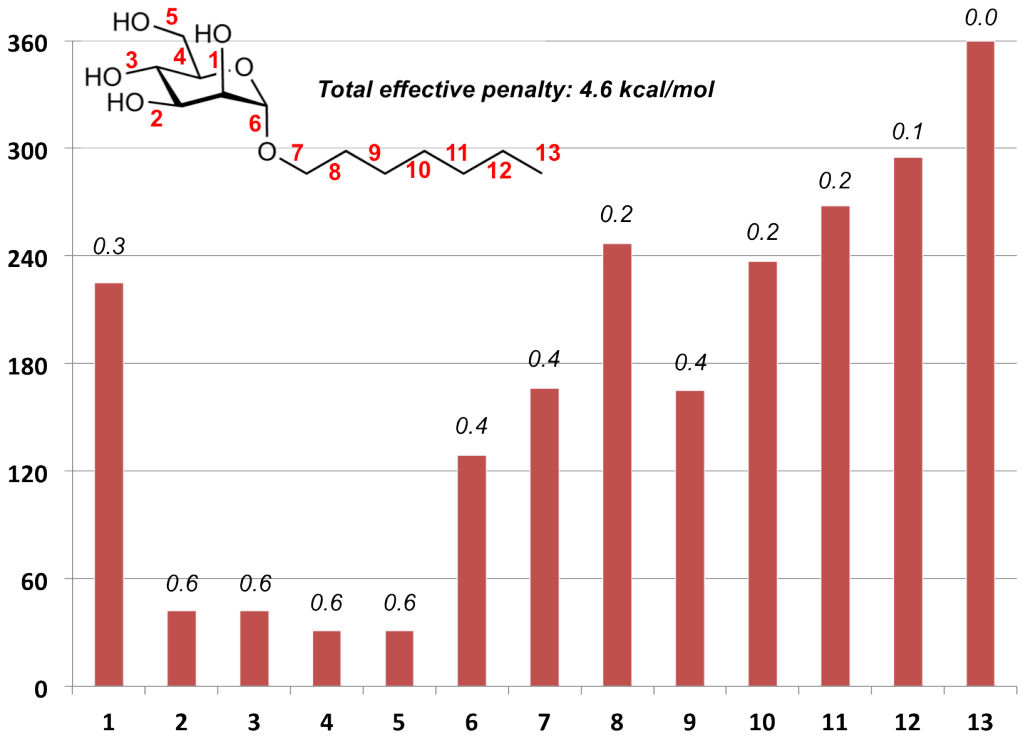


Figure 3.3: Result of the bond scanning procedure for heptyl α-D-mannoside binding to the FimH LD. The horizontal axis corresponds to the identified rotatable bonds (red labels), with their respective penalties [kcal/mol] given on top of each bar. The vertical axis corresponds to the rotation – in degrees – accessible to each bond.

4. Results and discussion

Layout

The following chapter summarizes the design and evaluation of selected groups of FimH inhibitors, performed based on the published biological and structural knowledge regarding the protein's lectin domain. The various parts of this work were done in collaboration with Prof. Beat Ernst, Dr. Said Rabbani (competitive binding assay measurements), Dr. Roland Preston (ITC measurements), Dr. Katharina Mayer, Dr. Katrin Lemme, Dr. Oliver Schwardt, Lijuan Pang, and Wojciech Schönemann. Next, the chapter proceeds to the development of the *VirtualDesignLab* framework, that automates and simplifies the design tasks for the convenience of medicinal chemists. This work was done with the help of Prof. Angelo Vedani, Dr. Martin Smieško, and Dr. Sameh Eid. Finally, attempts to model microsecond-scale behavior of the complete, two-domain FimH protein – based on advanced molecular dynamics protocols – are discussed, in the attempt of elucidating the conformational characteristics of the system under flow and static conditions. This part of the work was carried out entirely by me.

4.1 FimH inhibitor design and optimization

4.1.1 Analysis of known structures

Owing to the effort of our and other groups,[97,145] the past two decades of research resulted in the discovery of various mannosidic compounds displaying a high affinity toward the FimH protein. These studies revealed, that excellent potencies for the lectin domain (LD) of this target could be achieved by means of monovalent α-D mannosides mimicking the natural high-mannose epitopes expressed on the urothelial cells. More recently, further improvements could be accomplished by focusing on interactions between the aglycones of these carbohydrate derivatives and the so-called FimH *tyrosine gate* lined by protein residues Tyr48 and Tyr137 [Figure 4.1]. These findings are illustrated by several crystal structures of the FimH protein with various ligands, including a natural, tri-mannosidic motif [Table 2.3]. Inspection of these structures showed that all but one (corresponding to a compressed conformation discussed in Section 4.2.4) maintained a remarkably conserved LD and binding pocket, with the position of Tyr48 (responding to a specific ligand) proving the only significant variable. Depending on the respective arrangements of the aglycone and tyrosine residues, the so-called *in* and *out* binding modes (at times referred to as *open* and *closed*) are commonly addressed throughout literature[146]. In order to build on these findings, the characteristics of *in*- and *out*-binding needed to be assessed. To this end, two FimH LD crystal structures – with ligands bearing an *in*-bound butyl (PDB code 1UWF; resolution 1.7 Å) and an *out*-bound biphenyl aglycone (PDB code 3MCY; resolution 2.9 Å) – were processed and analyzed based on MD simulations and subsequent free-energy calculations.

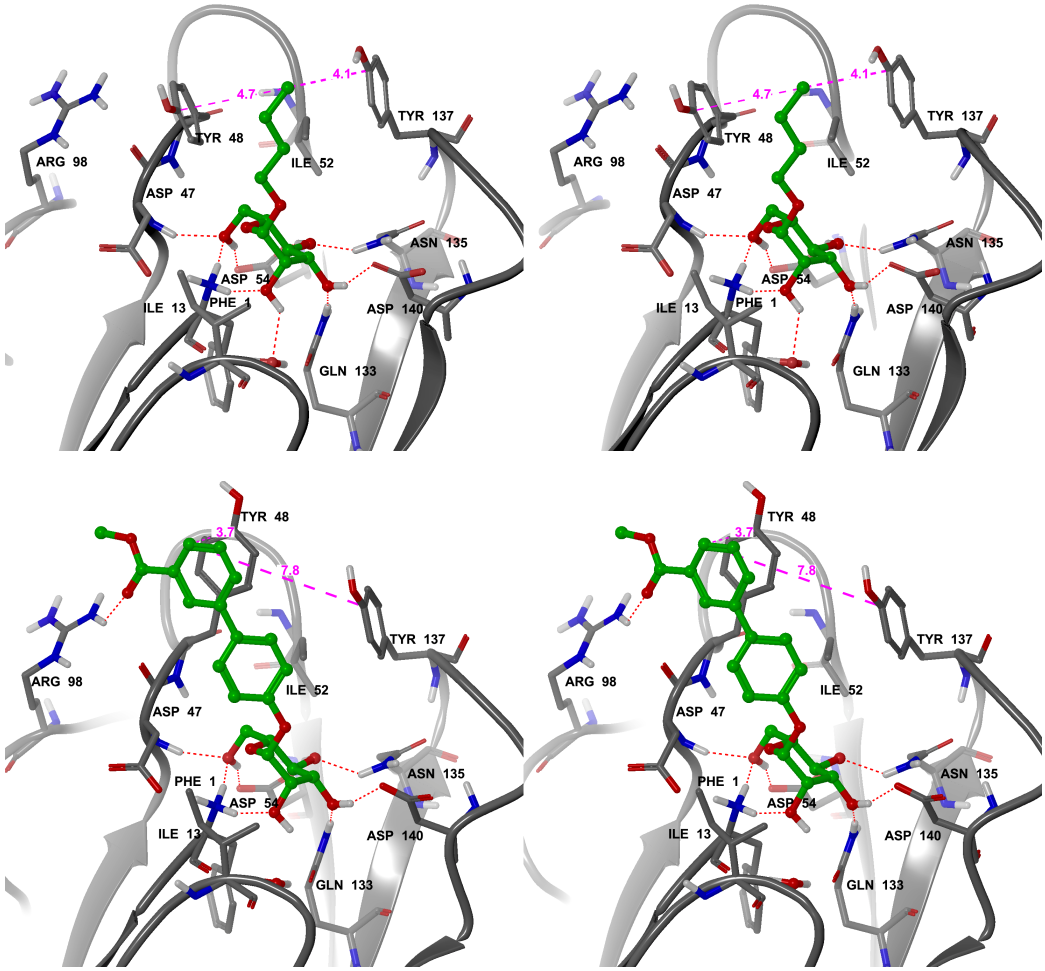


Figure 4.1: Stereo representations of butyl (top) and biphenyl (bottom) α-D-mannoside binding poses as obtained from computationally-refined 1UWF and 3MCY crystal structures. In the case of 1UWF, the refinement process led to changing the tyrosine gate alignment from *closed* to *open*.

Following structure preparation and refinement, interactive inspection of the complexes showed the presence of multiple hydrogen bonds around the mannose moiety along with hydrophobic interactions between the aglycones and the *tyrosine gate* [Figure 4.1]. The respective ligand poses have also proven to be stable throughout 2 ns MD simulations, confirming the preferences of either *in*- or *out*-binding depending on aglycone shape, size, and rigidity. MM/GBSA free-energy decomposition [Table 4.1] performed on the two trajectories revealed the particularly important role of residues Phe1, Asp47, Asp54, Gln133, Asn135, and Asp140 within the mannose-binding

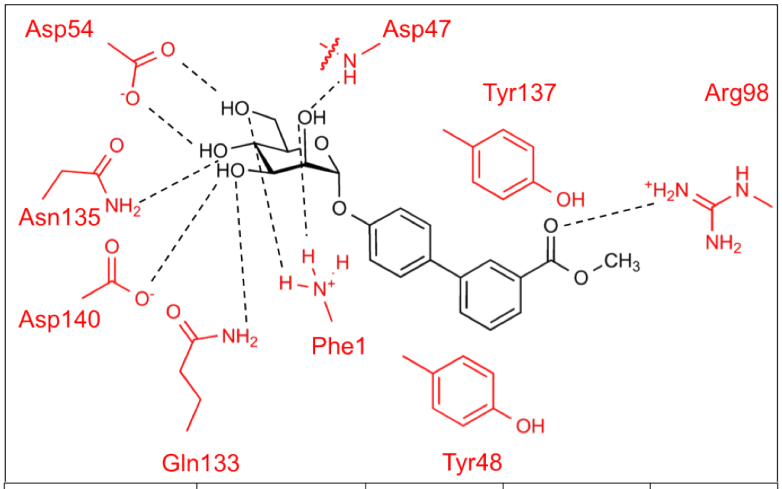
pocket. As also previously reported,[147] these residues (along with Ile13 and Ile52 engaged in hydrophobic interactions) were likely responsible for the bulk of affinity and specificity toward mannose-bearing ligands due to the formation of a tight hydrogen-bond network. Among them, Phe1 and Asp54 were the strongest contributors to the binding energy (around 5 kcal/mol each), due to each forming two very stable (present in the majority of all simulation frames), charge-assisted hydrogen bonds. It is assumed, that a lack of interaction with either one of these residues (i.e. caused by modification of the mannose hydroxyl groups) would lead to a complete loss of affinity. Altogether, a total of 8 stable hydrogen bonds (as opposed to previously reported 10) could be identified around the mannose moiety [Figure 4.2],[93] with a potential 9^{th} mediated by a conserved water molecule deep within the binding pocket (discussed in Section 4.1.4). An additional bond, reported to form between the *meta*-positioned methyl-ester substituent of the biphenyl aglycone and the residue Arg98 outside of the pocket,[97] could only be observed in a fraction of the MD frames (around 30%) and was considered of low significance. The discrepancy between modeling and experiment regarding the number of hydrogen bonds could be attributed to the *flash-cooling* procedures accompanying the latter. Specifically, the introduction of very low temperature (commonly 100 K), was likely to diminish some of the residual motion, allowing interactions that would, under physiological conditions (over 300 K), prove entropically unfavorable. Furthermore, as already highlighted in the case of the 1UWF structure,[94] the exposure of the binding interface on the protein surface could have led to crystal packing stabilization of less favorable arrangements. These assumptions could be validated by subjecting the 3MCY structure to an MD simulation in 100 K, that showed minimal amount of motion within the binding interface and excellent stability of all interactions, including the hydrogen bond with Arg98.

Upon ligand removal, 300 K MD simulations of the apo proteins were also performed in order to observe the behavior of the unoccupied binding pockets and their vicinity. Subsequent inspection indicated that while the site interior was preorganized and did not experience change upon mannose binding, the residues Tyr48 and Tyr137 maintained constant motion with the former sampling both *in* and *out* conformations. The prevalent Tyr48 side chain conformations were found in good agreement with the two most common tyrosine rotamers (corresponding to χ_1 angles of 66 and 180 degrees) found in the Schrödinger software library,[104] indicating that the mode preferences were influenced by the internal energy of this amino acid. Finally, clustering of the trajectory snapshots, shoved the *in* arrangement more often populated (both simulations), as also indicated by the available apo protein structures.[92]

Table 4.1: Selected per-residue contributions to the binding free energy of butyl- and biphenyl α-D-mannosides, obtained as average values from MD-based MM/GBSA (software *Amber*) calculations based on the respective crystal structures. Due to the employed implicit solvent model, the results are likely to be overestimated and serve only to highlight the most important residues. All values are given in kcal/mol.

Residue	**Butyl mannoside**	**Biphenyl mannoside**
Asp54	−5.7	−6.2
Phe1	−5.8	−4.3
Tyr48	−1.4	−3.3
Asn135	−2.4	−2.6
Asp47	−1.9	−2.1
Ile52	−1.5	−2.0
Gln133	−2.7	−1.9
Arg98	0.0	−1.8
Asn46	−1.4	−1.4
Asp140	−0.4	−1.1
Ile13	−1.5	−0.8
Phe142	−0.5	−0.5
Tyr137	−0.4	−0.3
His45	−0.3	−0.3
Thr51	0.0	−0.2
Asn138	−0.3	−0.2
Thr53	−0.1	−0.1
Phe144	−0.1	−0.1
Cys44	−0.1	−0.1

Based on these findings, it was concluded that further affinity improvements for monovalent mannosidic ligands could be obtained either via 1) aglycone alterations maximizing the favorable contacts with the *tyrosine gate* or, 2) modifications of the mannose moiety introducing new interactions while retaining the existing ones.



Acceptor atom	Donor atom	Fraction of frames	Avg. distance [Å]	Avg. angle
Ligand: O5	Asp47: N	0.56	2.90	160
Ligand: O4	Asn135: ND2	0.55	2.87	156
Ligand: O5	Phe1: N	0.70	2.86	154
Ligand: O3	Gln133: NE2	0.36	2.91	159
Ligand: O7	Arg98: NH2	0.33	2.85	148
Ligand: O2	Phe1: N	0.41	2.88	152
Asp54: OD2	Ligand: O5	1.00	2.56	167
Asp54: OD1	Ligand: O4	1.00	2.62	167
Asp140: OD1	Ligand: O3	0.96	2.69	165

Figure 4.2: (top) Two-dimensional representation of all hydrogen bonds identified around the biphenyl ligand of the 3MCY crystal structure, along with (bottom) their frequencies and geometries throughout a MD simulation. Unless stated otherwise, the strongest among these bonds were also found for all other compounds presented in this work.

4.1.2 Evaluation protocols

Due to the substantial number of screened structures (close to 150), MD simulations and post-processing analyses for all of them could not be accomplished within the scope of this thesis. Hence, the majority of evaluations was performed based on interactive inspection (focusing on aglycone orientation and hydrogen bonding) and scoring of single docking structures, with only selected cases subjected to a more through MD-based analysis. To this end, unified and reliable means of ligand evaluation had to be established. For the purposes of this work, scoring using the well-established *Glide SP* function implemented in the *Induced Fit Docking* protocol along with MM/GBSA post-processing methods (software *Prime* for docking and *Amber* for MD analyses, unless stated otherwise) were selected. Given the possibility of bias

toward larger ligands,[148] an additional efficiency metric defined by dividing the MM/GBSA scores by the number of heavy ligand atoms was also established. In addition, representative compounds in each group were subjected to the *Cheetah* and *Aquarius* technologies,[44,134] employing an explicit solvent representation and the directional *Yeti* force-field.[52] Applying these protocols toward re-docking the two compounds discussed in the previous section to their corresponding protein grids (i.e. refined experimental conformations), resulted in top-scored poses within an RMSD of 1 Å from the crystal structures, with all crucial interactions reproduced and the biphenyl ligand correctly ranked above the alkyl derivative. Cross-docking (i.e. switching the protein grids) of the molecules was also successful, with the correct Tyr48 rotamers re-assigned in both cases. Subsequent MD simulations with the docked poses could prove them stable, showing no deviation from the experimental structures. Altogether, the chosen computational protocols were deemed capable of reliably docking and ranking prospective compounds and accounting for the necessary protein flexibility. Still, it should be emphasized that the properties of the underling implicit solvent models and lack of entropy account, have likely led to an overestimation of the MM/GBSA scores.[149] Thus, the presented work was primarily meant to diminish the pool of synthesis candidates, rather than yield accurate predictions. Later on, a larger set of compounds was used to derive approximate scaling coefficients for the MM/GBSA results – 1/10 for *Prime* and 1/7 for *Amber* [Appendix Figure 7.1] –, and to establish a vastly more accurate QSAR model (discussed in Section 4.2).

Further discussion in this chapter will focus on a selected subset of promising compounds, with special emphasis on those presented in already published work. For every group discussed, a single representative compound is highlighted along with its detailed interaction profile obtained from an MD simulation. The entire sets of tested derivatives are listed in Tables 7.1 and 7.2 of the Appendix along with their docking, MM/GBSA, and ligand efficiency scores.

4.1.3 Aglycone modifications

Various biphenyl derivatives introduced around the time this work began, have proven among the most potent FimH LD binders.[150] Based on available crystal structure knowledge, it could be derived that their excellent affinities were related to π-π stacking interactions between the residue Tyr48 and the second ("outer") phenyl ring of the their aglycones. In addition, the free-energy decomposition of the biphenyl compound presented in section 4.1.1, indicated a significantly weaker interaction with the second "gate" residue

– Tyr137. This was attributed to a too high rigidity and bulkiness of the aglycone, preventing positioning between the tyrosine side chains and forcing them into a non-native, perpendicular orientation (a general characteristic of the *out*-binding mode). Hence, attempts were made to either insert larger aglycones between the parallel-aligned *tyrosine gate* residues (while retaining the favorable stacking interactions), or modify the biphenyl scaffold to optimize the *out*-binding pattern. To this end, series of compounds containing various aglycone modifications were screened *in silico* against the LD structure. In each case, multiple residue/ligand conformations – generated by the flexible docking protocol – were tested in order to account for individual mode preferences. In summary, the majority of the screened compounds could be classified between the following groups (cf. Appendix Table 7.1):

- Biphenyl compounds with various substitution patterns on one or both rings
- Structures introducing linkers before and/or between the aromatic moieties
- Squaric acid derivatives and similar compounds, altering the shape of the biphenyl aglycone by introducing a unique ring system
- Triazole and aminocyclobutanol derivatives, exploring the potential of replacing the first ("inner") phenyl ring
- Indolyl, indolinyl, and pyrrolopyridine derivatives replacing the outer phenyl ring with a larger system
- Derivatives with saturated or heteroaromatic ring systems
- Compounds bearing alkyl aglycones, primarily used for reference purposes

Optimization of the biphenyl substitution pattern

Following the discovery of the inhibitors based on a rigid biphenyl scaffold, it could also be observed that the introduction of aromatic ring substituents could lead to further gains in potency. Initial work in this direction,[145] revealed ligands with *ortho*-chloro and *para*-carboxyl groups on inner and outer rings respectively, to reach *in vitro* affinities in the low-nanomolar ranges. Thus, further optimization attempts in this direction were undertaken [Appendix Table 7.1: 1–21].

Flexible docking of a representative compound bearing the both above-mentioned substituents, resulted in a higher scored *out*-binding pose, closely resembling the one of the previously discussed co-crystallized 3MCY ligand [Figure 4.3]. Subsequent MD simulations indicated a high stability of this pose, allowing excellent stacking interactions with Tyr48 (2.4 kcal/mol). This further confirmed that the modest resolution (2.9 Å) of the experimental structure and the crystallization conditions did not lead to an invalid representation of biphenyl ligand binding. Still, a somewhat higher residual mobility of our compound could be noted (likely to the benefit of conformational entropy), owing to the outer-ring substituent not forming directional interactions with the protein. It could thus be stated, that the *out*-binding mode was related to the large size, linear shape, and rigidity of the aglycones rather than the presence of specific interactions or conditions. With respect to the halogen substituent, it occupied a niche between the Tyr137 side chain and the mannose binding pocket. Given the predominately hydrophobic character of this cavity and lack of suitable acceptor groups, a possibility of a chlorine-specific interaction (e.g. a halogen bond) was excluded and the increased affinity attributed to van der Waals contacts of the atom. Subsequent synthesis and evaluation of compounds with the chlorine replaced by other moieties have validated this, showing potency improvements with increasing substituent van der Waals volumes. Nevertheless, deviating from a symmetric character of the groups (introduction of a metoxy and cyclopropyl substituents) resulted in the deterioration of the *in vitro* potencies [Figure 4.3]. Based on the analysis of docking structures it could be concluded, that the niche preferred groups of spherical shape and/or that the entropic component related to restricting the rotatable, non-symmetric groups, diminished the benefit of the added interactions. As a result, a compound with a trifluoromethyl group replacing the chlorine atom was synthesized and has proven the most potent in the series [Table 4.2]. It is likely that this moiety corresponds to the maximum substituent size that can be accommodated without distorting the optimal position of the scaffold atoms. This work has been reported as part of a recent publication.[151]

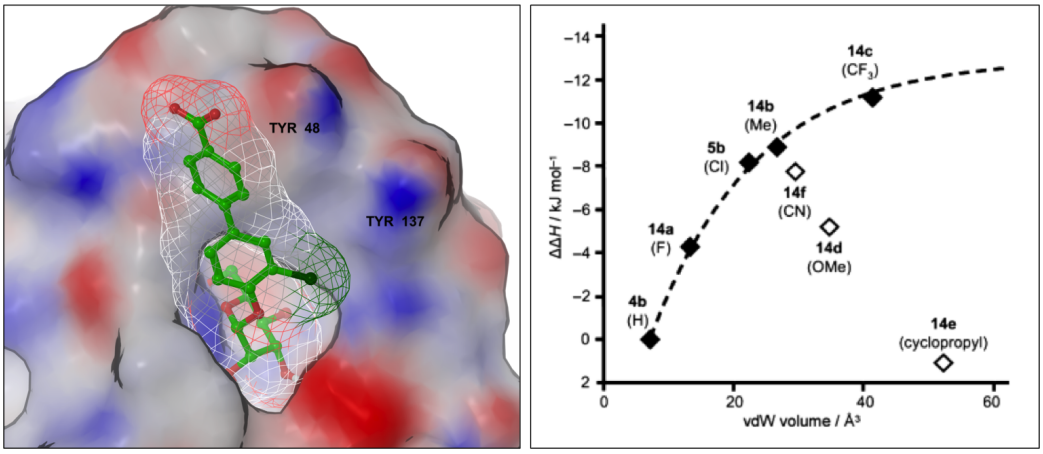
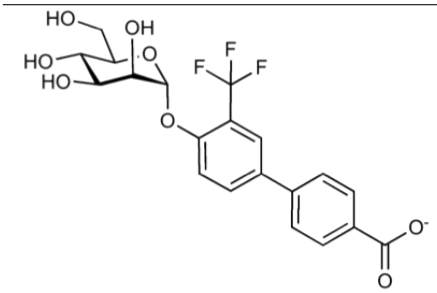


Figure 4.3: (left) Binding pose of a representative biphenyl α-D-mannoside, bearing an *ortho*-chloro substitution on its inner phenyl ring. Surface coloring corresponds to hydrophobic (gray), positively-charged (blue), and negatively-charged (red) regions. (right) Relationship between experimental binding enthalpies and van der Waals volumes of various *ortho* aglycone substituents. Figure adapted from [151].

MD simulations for various compounds aimed at optimizing the substitution pattern of the outer phenyl ring were also performed. Up till date however, the *para*-positioned carboxyl/methyl-ester groups remain among the most beneficial. In all likelihood, the presence of these moieties (with their strong electron-withdrawing properties) favorably influences the π-π stacking with the electron-rich Tyr48 side chain, while increased desolvation costs are avoided through a high solvent-exposure of the aglycone. This, in turn, has led to a conclusion that groups with stronger electron-withdrawing properties could prove even more advantageous.

Various other biphenyl compounds – introducing linker groups or altering the linear phenyl ring arrangement – screened as part of this study, have proven inferior to the above-discussed derivatives. Given the good *in silico* scores for some of them, the weaker affinities were attributed to entropy and solvation-related components of their binding – aspects difficult to properly address with the employed methods.

Table 4.2: Binding characteristics of a reference biphenyl derivative, including constituents of the binding free energy (top; scaling coefficient 1/7), per-residue free energy contributions (right), and most stable hydrogen bonds (bottom). All results obtained either using *Amber* (averages over MD trajectories) or *Cheetah* and *Aquarius* software (docking pose; values in brackets) and given in kcal/mol and Å. The relatively small desolvation costs are a consequence of a high solvent-exposure of the bound ligand (applies also for further results).



Energy	Contribution	Residue	Contribution
van der Waals	−24.1 (−20.9)	Phe1	−4.8
electrostatic	−59.5 (−41.9)	Ile13	−1.0
solvation	+25.2 (+32.1)	Cys44	−0.1
H-bond	n/a (−24.1)	His45	−0.3
total	−58.4 (−54.8)	Asn46	−1.5
		Asp47	−2.4
Exp. affinity	8.9 nM	Tyr48	−2.4
		Thr51	−0.2
		Ile52	−2.0
		Thr53	−0.1
		Asp54	−5.9

Acceptor atom	Donor atom	% frames	Distance	Angle
Ligand: O4	Asn135: ND2	60	2.9	156
Ligand: O5	Asp47: N	60	2.9	161
Ligand: O3	Gln133: NE2	40	2.9	162
Ligand: O5	Phe1: N	70	2.9	151
Ligand: O2	Phe1: N	60	2.9	152
Asp54: OD2	Ligand: O5	100	2.6	168
Asp54: OD1	Ligand: O4	100	2.6	168
Asp140: OD2	Ligand: O3	100	2.7	165

Arg98	−0.1
Gln133	−2.2
Asn135	−2.6
Tyr137	−0.4
Asn138	−0.2
Asp140	−0.6
Phe142	−0.2
Phe144	−0.1

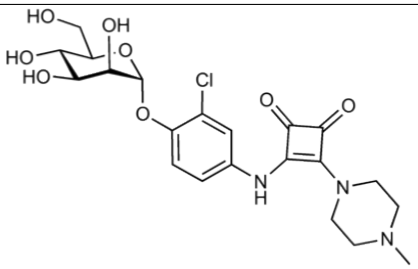
Squaric acid derivatives

Given the vicinity of polar groups (including the side chains of residues Tyr48, Tyr137, and Thr51), enhancing the ligands with complementary moieties remained a vital design approach. Still, previous biphenyl derivatives have proven incapable of taking advantage of these groups, owing to the rigidity and linearity of their molecular scaffolds. Thus, a subset of compounds [Appendix Table 7.1: 22–27] with a squaric acid moiety altering both the shape and the flexibility of the aglycones was screened *in silico*. A secondary objective regarding these compounds was exploring the possibility of stacking between the *tyrosine gate* and alternative, non-phenyl ring systems. According to previous reports,[152] compounds of this type could also potentially form covalent bonds between the squaric acid and the charged N-terminus

of residue Phe1 in the mannose-binding pocket. These considerations were however discarded based on a more recent publication as well as experimental affinities below of what one would expect form the formation of a covalent linkage.[153]

In line with our intentions, docking poses obtained for the compounds were often accompanied by hydrogen-bond formation between the oxygen atoms of the added cycle and the tyrosine residues. Countering these indications though, further analyses based on MD simulations, did not confirm the stability of the interactions [Table 4.3]. This was – yet again – likely due to an entropic penalty that would have to be paid for restraining the aglycone upon forming the bonds. Nevertheless, the unique, curved shape enforced by the added group – itself successfully taking over the role of stacking with Tyr48, as apparent from the energy decomposition – enabled the distal ends of the aglycones to form additional contacts with the vicinity of the previously mentioned Arg98. Interestingly, while the subsequent biological evaluation of the synthesized compounds did not prove them superior to the most potent LD binders (affinities in the mid- to low-nanomolar ranges), some of them exhibited similar potencies across a variety of assays and FimH variants.[154] This gave reason to conclude that the compounds could either bind equally well to multiple FimH conformational states (discussed elsewhere in this thesis) or enforce one of these states independently from the conditions. Work on this subject is currently undergoing and has recently resulted in a crystal structure of one such compound with the FimH LD (personal communication – Dr. Roland Preston). The corresponding binding mode was found in good agreement with the computationally-derived pose [Figure 4.4] and has also confirmed (through the presence of multiple ligand conformations) that the previously assumed high residual mobility of the aglycone was indeed present.

Table 4.3: Binding characteristics of a representative squaric acid derivative, including constituents of the free energy (top; scaling coefficient 1/7), per-residue energy contributions (right), and most stable hydrogen bonds (bottom). Results obtained either using *Amber* (averages over MD trajectories) or *Cheetah* and *Aquarius* software (docking pose; values in brackets) and given in kcal/mol and Å.



Energy	**Contribution**	**Residue**	**Contribution**
van der Waals	−29.8 (−16.7)	Phe1	−5.6
electrostatic	−91.7 (−43.5)	Ile13	−1.4
solvation	+54.2 (+25.7)	Cys44	−0.1
H-bond	n/a (−26.0)	His45	−0.4
total	−67.3 (−60.5)	Asn46	−1.5
		Asp47	−2.9
Exp. affinity	21.1 nM	Tyr48	−3.9
		Thr51	−0.2
		Ile52	−1.9
		Thr53	−0.1

Acceptor atom	**Donor atom**	**% frames**	**Distance**	**Angle**	Asp54	−6.1
Ligand: O4	Asn135: ND2	60	2.9	158	Arg98	−0.3
Ligand: O5	Asp47: N	60	2.9	161	Gln133	−2.4
Ligand: O3	Gln133: NE2	50	2.9	162	Asn135	−2.6
Ligand: O5	Phe1: N	80	2.9	153	Tyr137	−0.4
Ligand: O2	Phe1: N	70	2.9	152	Asn138	−0.2
Asp54: OD2	Ligand: O5	100	2.6	168	Asp140	−0.7
Asp54: OD1	Ligand: O4	100	2.6	168	Phe142	−0.3
Asp140: OD2	Ligand: O3	100	2.7	166	Phe144	−0.1

On a separate note, some of the compounds in this series showed that aglycones substantially longer than the biphenylic ones, yielded little to no affinity improvement. A limitation like this has already been indicated for the previously introduced alkyl α-D-mannosides,[94] with ligands longer than the heptyl derivative not showing improved potencies due to exposing hydrophobic parts of the aglycone to the solvent without gaining further interactions. Regarding the compounds discussed in this work, it could be derived that aglycone lengths exceeding 10-11 Å (measured as the distance between the anomeric oxygen and the furthest aglycone atom) were unlikely to benefit potency. Attaching specific groups beyond this point could however favorably affect PK properties at a acceptable expense of affinity. It should also be noted that, both the initial docking and subsequent MD simulations for some of the compounds, hinted on their ability to adapt/switch between poses that were often difficult to classify in terms of *in*/*out* binding. This suggests that categorizing between two arbitrary docking modes, could prove misleading.

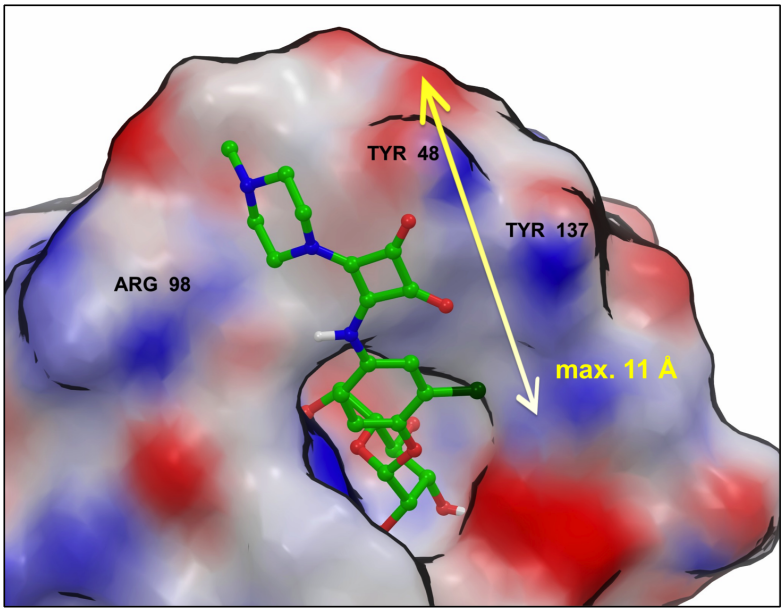


Figure 4.4: Docking pose of a representative squaric acid derivative. Yellow arrow illustrates an approximate aglycone length beyond which no interactions with the protein may be formed. Surface coloring corresponds to hydrophobic (gray), positively-charged (blue), and negatively-charged (red) regions.

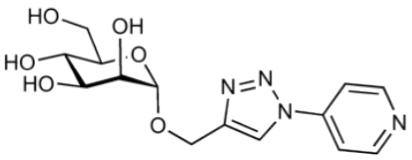
Triazole derivatives

Based on the assumption, that positioning the ligand aglycone between the two parallel-aligned tyrosine residues should benefit the affinities – as already proven for inhibitors with alkyl moieties –,[155] a series of triazole derivatives was screened *in silico*, synthesized, and tested *in-vitro*.[146] The design of these derivatives was also related to the observation that the direction in which the biphenyl aglycone was oriented (as seen in the 3MCY crystal structure), differed from that of alkyl α-D-mannosides [Figure 4.1]. This in turn could be attributed to the geometry imposed by the tightly bound, inner phenyl ring, lacking rotatable bonds around it. This cycle was thus substituted with a smaller triazole system preceded by a flexible linker group [Appendix Table 7.1: 28–37].

Docking and MD simulations performed for compounds in this group, indicated successful reorientation of the aglycones into an *open* tyrosine gate [Figure 4.6]. Later on this could be – to an extent – confirmed by the emergence of crystal structures with similar compounds bearing flexible linker groups alongside bulky ring systems.[147] Opposing the initial assumptions (though in line with the generally weaker docking scores), the *in vitro* affinities for the series, have proven inferior to the most potent biphenyl and alkyl

inhibitors. A more detailed inspection of the corresponding MD trajectories hinted on diminished mobility of the protein-ligand interfaces with respect to alkyl α-D-mannosides. This suggested that the entropic penalty imposed on the natively mobile but also bulky triazole-bearing aglycone and the *tyrosine gate*, overcompensated the already modest enthalpic gains. [Table 4.4]. As such, it could be concluded, that only compounds with slim, flexible aglycones (e.g. heptyl α-D-mannoside that perfectly complements the shape and capacity of the *tyrosine gate*) could fully benefit from an *in*-binding mode, whereas larger scaffolds should be kept rigid and prioritize interactions with residues accessible within the *out* mode. Still, this class of inhibitors remains noteworthy due to their excellent aqueous solubilities – a prerequisite toward oral bioavailability.

Table 4.4: Binding characteristics of a representative triazole derivative, including constituents of the binding free energy (top; scaling coefficient 1/7), per-residue free energy contributions (right), and most stable hydrogen bonds (bottom). All results obtained either using *Amber* (averages over MD trajectories) or *Cheetah* and *Aquarius* software (docking pose; values in brackets) and given in kcal/mol and Å.



Energy	**Contribution**	**Residue**	**Contribution**
van der Waals	−25.3 (−18.1)	Phe1	−3.5
electrostatic	−76.9 (−41.9)	Ile13	−0.9
solvation	+46.5 (+22.0)	Cys44	−0.1
H-bond	n/a (−27.7)	His45	−0.3
total	−55.8 (−65.7)	Asn46	−1.4
		Asp47	−1.9
Exp. affinity	70.0 nM	Tyr48	−3.0
		Thr51	−0.5
		Ile52	−2.2
		Thr53	−0.1
		Asp54	−6.2

Acceptor atom	**Donor atom**	**% frames**	**Distance**	**Angle**	Arg98	0.0
Ligand: O5	Phe1: N	70	2.8	150	Gln133	−1.6
Ligand: O4	Asn135: ND2	70	2.9	158	Asn135	−2.6
Ligand: O2	Phe1: N	60	2.9	150	Tyr137	−1.0
Ligand: O5	Asp47: N	40	2.9	159	Asn138	0.0
Ligand: O3	Gln133: NE2	30	2.9	161	Asp140	−0.1
Asp54: OD2	Ligand: O5	100	2.6	168	Phe142	−0.6
Asp54: OD1	Ligand: O4	100	2.6	168	Phe144	−0.1

On a side note, NMR experiments performed as part of our published study, revealed that the removal of the anomeric oxygen and attachment of the aglycones directly to their mannose moieties, could result in an unusual $^1C_4$ chair conformation of the latter [Figure 4.5]. Due to substantial loss of compound affinities, this structural element is thus considered crucial and should not be tampered with.

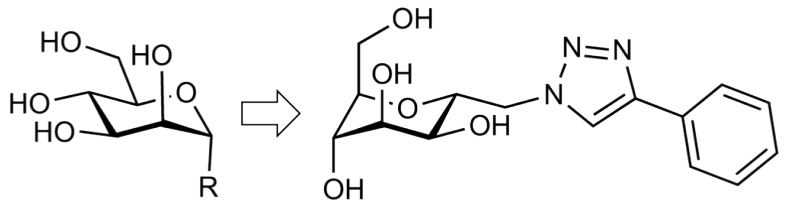


Figure 4.5: Illustration of the unfavorable $^1C_4$ mannose flip discovered – by means of ^{1}H NMR experiments – for some for the presented (right) and additional compounds in the triazole series.

Indolyl, indolinyl and pyrrolopiridine derivatives

As indicated with the squaric acid derivatives, the outer phenyl ring of the biphenyl scaffold could be substituted without overly diminishing the potencies of the compounds. A further attempt in this direction, involved the replacement of the ring with indolyl and indolinyl moieties [Appendix Table 7.1: 38–54]. Here, a secondary goal was the improvement of the generally poor aqueous solubility of previous compounds, by means of distorting the symmetry and planarity of the aglycone. Strategies like this, have been previously successfully applied toward optimizing the properties of various compounds.[156]

In silico evaluation of representative compounds, indicated some of them as excellent LD binders (though their scores were possibly biased by the larger size of the aglycones as indicated by the ligand efficiency data). The primary reason for this, was the increased contact surface offered by the fused-ring systems, enabling more interactions while otherwise replicating all the favorable contacts of the biphenyl ligands [Table 4.5]. Furthermore, these compounds could fully benefit form the ring substitution effects (i.e. *ortho* substitution on the inner ring along with an electron withdrawing moiety on the outer) discovered for the biphenyl series. Still, the most soluble derivatives (generally bearing the least planar aglycones) proved incapable of equally good interactions, resulting in lower scores. These observations were found in good agreement with a later biological evaluation in which indolyl compounds with 30-45 degree angles between respective ring systems, were

found most potent.[157] These derivatives are in fact among the most potent FimH inhibitors till date, rivaling the *in vivo* efficacy of the standard UTI antibiotic treatment – ciprofloxacin. Also, an additional confirmation toward the validity of the modeling could be gained through a recently resolved crystal structure (personal communication – Dr. Roland Preston), showing an orientation of an indolyl derivative that was nearly indistinguishable form its docking pose [Figure 4.6].

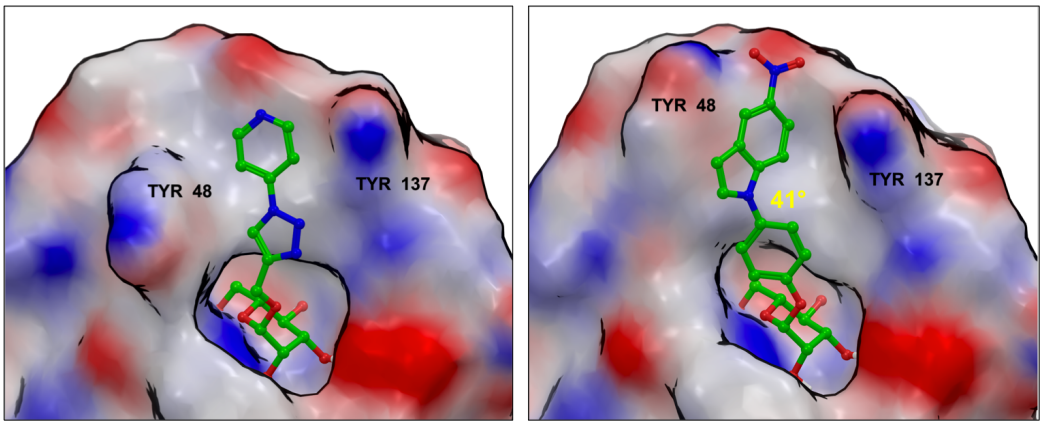
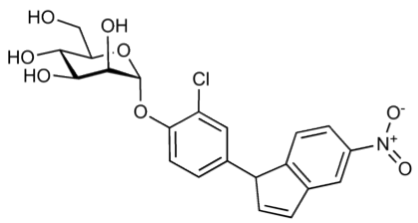


Figure 4.6: (left) Predicted binding pose of the most potent and best-scored of the triazole inhibitor series. (right) Docking pose of a representative indolyl derivative. The yellow number corresponds to the dihedral angle between respective ring systems of the compound. Surface coloring corresponds to hydrophobic (gray), positively-charged (blue), and negatively-charged (red) regions.

Table 4.5: Binding characteristics of a representative indole derivative, including constituents of the binding free energy (top; scaling coefficient 1/7), per-residue free energy contributions (right), and most stable hydrogen bonds (bottom). All results obtained either using *Amber* (averages over MD trajectories) or *Cheetah* and *Aquarius* software (docking pose; values in brackets) and given in kcal/mol and Å.



Energy	**Contribution**	**Residue**	**Contribution**
van der Waals	−26.2 (−16.0)	Phe1	−4.3
electrostatic	−83.4 (−35.9)	Ile13	−1.3
solvation	+48.8 (+20.2)	Cys44	−0.1
H-bond	n/a (−16.2)	His45	−0.4
total	−60.7 (−47.9)	Asn46	−1.1
		Asp47	−2.2
Exp. affinity	7.7 nM	Tyr48	−2.5
		Thr51	−0.3
		Ile52	−2.2
		Thr53	−0.1

Acceptor atom	**Donor atom**	**% frames**	**Distance**	**Angle**	Asp54	−7.1
Ligand: O5	Phe1: N	90	2.9	154	Arg98	0.0
Ligand: O4	Asn135: ND2	70	2.9	160	Gln133	−2.0
Ligand: O2	Phe1: N	70	2.9	153	Asn135	−2.1
Ligand: O3	Gln133: NE2	60	2.9	163	Tyr137	−0.6
Ligand: O5	Asp47: N	30	2.9	163	Asn138	−0.1
Asp54: OD2	Ligand: O5	100	2.6	167	Asp140	−0.8
Asp54: OD1	Ligand: O4	100	2.6	168	Phe142	−0.4
Asp140: OD1	Ligand: O3	70	2.7	165	Phe144	−0.2

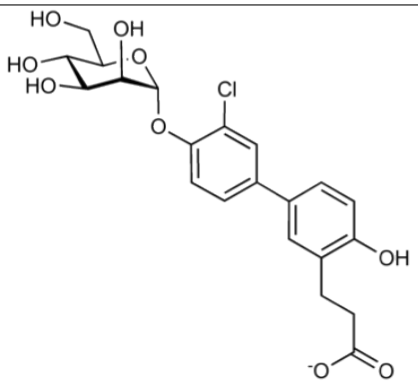
Other compounds of interest

Two additional groups of compounds were screened *in silico* but await synthesis and biological evaluation [Table 7.1: 55-62].

The first was a small set of derivatives bearing carboxyl and hydroxyl groups flexibly-linked to the distal end of their aglycones [Figure 4.7]. The primary aim of these compounds was the improvement of potential bioavailability by enabling them to form lactones when not bound to their protein target.[158,159] Through the span of the analysis, it could additionally be revealed that two of the tested structures, bearing carboxyl groups linked to the *meta*-phenyl positions, formed stable salt bridges to the Arg98, resulting in *in silico* scores superior to previous biphenyl compounds [Table 4.6]. It is likely that the long, flexible linkers introduced as part of the substituents, enabled an interaction that could not be maintained by the ligand of the

3MCY crystal structure. Also, though tested with a biphenyl scaffold, the favorable properties of the added groups could likely be retained with other aglycones, as long as they preserved a similar shape and size.

Table 4.6: Binding characteristics of a representative lactone-forming derivative, including constituents of the binding free energy (top; scaling coefficient 1/7), per-residue free energy contributions (right), and most stable hydrogen bonds (bottom). All results obtained either using *Amber* (averages over MD trajectories) or *Cheetah* and *Aquarius* software (docking pose; values in brackets) and given in kcal/mol and Å.



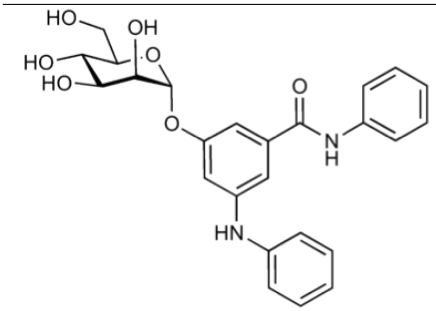
Energy	Contribution	Residue	Contribution
van der Waals	−25.8 (−19.8)	Phe1	−5.9
electrostatic	−70.6 (−39.2)	Ile13	−1.1
solvation	+31.9 (+23.4)	Cys44	−0.1
H-bond	n/a (−12.9)	His45	−0.4
total	−64.5 (−47.5)	Asn46	−1.6
		Asp47	−2.5
Exp. affinity	no data	Tyr48	−3.1
		Thr51	−0.2
		Ile52	−1.9
		Thr53	−0.1
		Asp54	−6.0
		Arg98	−1.9
		Gln133	−2.4
		Asn135	−2.7
		Tyr137	−0.4
		Asn138	−0.2
		Asp140	−0.7
		Phe142	−0.2
		Phe144	−0.1

Acceptor atom	Donor atom	% frames	Distance	Angle
Ligand: O5	Phe1: N	80	2.9	153
Ligand: O2	Phe1: N	70	2.9	152
Ligand: O4	Asn135: ND2	60	2.9	156
Ligand: O5	Asp47: N	60	2.9	162
Ligand: O3	Gln133: NE2	50	2.9	163
Asp54: OD1	Ligand: O4	100	2.6	168
Asp54: OD2	Ligand: O5	100	2.6	168
Asp140: OD2	Ligand: O3	100	2.7	165

The second inhibitor group consisted of "bifurcated" compounds with two aromatic rings connected to a shared CH/phenyl core [Figure 4.7]. The idea here was to enable interactions common for both binding modes, by positioning one of the rings between the tyrosine residues and the other within the vicinity of Arg98. Molecular dynamics simulations showed the compounds designed around the phenyl core as capable of sustaining their excellent initial docking poses. The same could not be said for the remaining structures due to a too high scaffold flexibility. Energy calculations performed on the antagonists, indicated improvement of their affinities [Table 4.7] (though, given the compound sizes, the ligand efficiency scores have to be kept in

mind). If proven even comparably potent to previously-tested derivatives, further improvement of these scaffolds could easily be obtained by means of various aromatic substituents.

Table 4.7: Binding characteristics of a representative bifurcated derivative, including constituents of the binding free energy (top; scaling coefficient 1/7), per-residue free energy contributions (right), and most stable hydrogen bonds (bottom). All results obtained either using *Amber* (averages over MD trajectories) or *Cheetah* and *Aquarius* software (docking pose; values in brackets) and given in kcal/mol and Å.



Energy	**Contribution**	**Residue**	**Contribution**
vdW	−27.0 (−18.1)	Phe1	−5.3
electrostatic	−88.6 (−41.8)	Ile13	−1.3
solvation	+51.3 (+11.9)	Cys44	−0.1
H-bond	n/a (−24.9)	His45	−0.4
total	−64.2 (−72.9)	Asn46	−1.4
		Asp47	−2.8
Exp. affinity	no data	Tyr48	−3.1
		Thr51	−0.2
		Ile52	−1.9
		Thr53	−0.1
		Asp54	−6.9
		Arg98	−0.3
		Gln133	−2.2
		Asn135	−2.4
		Tyr137	−0.4
		Asn138	−0.4
		Asp140	−0.5
		Phe142	−0.3
		Phe144	−0.1

Acceptor atom	**Donor atom**	**% frames**	**Distance**	**Angle**
Ligand: O5	Phe1: N	80	2.9	152
Ligand: O2	Phe1: N	70	2.9	152
Ligand: O4	Asn135: ND2	60	2.9	158
Ligand: O5	Asp47: N	50	2.9	162
Ligand: O3	Gln133: NE2	50	2.9	163
Asp54: OD1	Ligand: O4	100	2.6	168
Asp54: OD2	Ligand: O5	100	2.6	168
Asp140: OD2	Ligand: O3	100	2.7	164

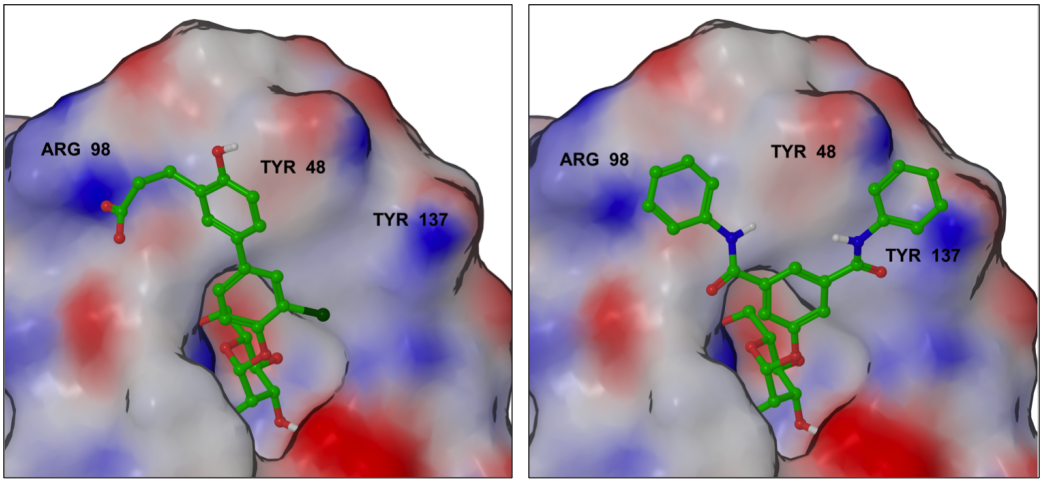


Figure 4.7: (left) Proposed binding pose of a representative lactone-forming compound, establishing a salt bridge with residue Arg98. (right) Docking pose of the top-scored "bifurcated" antagonist. Surface coloring corresponds to hydrophobic (gray), positively-charged (blue), and negatively-charged (red) regions.

Mutation study

The work on prospective FimH inhibitors revealed different residual contributions toward coordinating the various aglycone types. Attempts of elucidating these relationships however, have met with several inconsistencies between modeling and experiment. Furthermore, the native flexibility discovered for the *tyrosine gate* called for assessing its role in greater detail. Hence, an approach was devised to selectively mutate the two tyrosines into alanines. This way the impact of respective side chains would be negated, allowing subsequent biological evaluation to reveal the consequences on ligand binding. Following expression and purification of the mutated proteins, competitive binding assay and ITC measurements (with the latter providing valuable insight into the entropy changes) were employed to probe the potency of ligands representing the *in* and *out* binding modes. The obtained results however, proved difficult to interpret. Specifically, the data showed a significant loss of affinity upon Tyr137 mutation, with the lack of Tyr48 being of much lesser consequence (for *in* and *out* binding compounds alike). What made this discovery surprising, was the fact that stacking with Tyr48 (much more so than interactions with Tyr137) were already confirmed responsible for the improved potencies of biphenyl ligands (Section 4.1.1). Furthermore, no indication could previously be found as to this residue holding any less significance toward the *in*-mode binding of alkyl α-D-mannosides than Tyr137. To elucidate the structural consequences of the Y48A and Y137A mutations,

protein grids based on modified FimH LD structures were constructed, MD-refined, and employed alongside the wild type models.

Table 4.8: Experimental (ITC) end computational (*Prime* MM/GBSA scores averaged over MD trajectories) results obtained for selected protein-ligand systems. All values in kcal/mol.

Compound	Mutant	ΔG_{ITC}	ΔH_{ITC}	$-T\Delta S_{ITC}$	$\Delta G_{MM\text{-}GBSA}$	$SEM_{MM\text{-}GBSA}$*
1,5-Anhydro D-mannitol (no aglycone)	WT	-8.2	-8.7	0.5	-48.2	0.3
	Y48A	-7.6	-8.6	0.9	-48.4	0.2
	Y137A	-7.6	-8.4	0.8	-50.0	0.3
Heptyl α-D-mannoside	WT	-10.2	-9.5	-0.7	-65.9	0.4
	Y48A	-10.0	-8.0	-2.0	-60.7	0.4
	Y137A	-9.1	-6.9	-2.2	-60.8	0.4
Biphenyl α-D-mannoside	WT	-10.5	-10.6	0.1	-65.6	0.4
	Y48A	-10.2	-9.6	-0.6	-61.1	0.4
	Y137A	-9.6	-7.7	-1.9	-58.0	0.3

* Standard error of the mean

MM/GBSA binding free energy calculations performed on respective ensembles of ligand-bound MD snapshots confirmed that the mutation Y48A resulted in diminished protein-ligand interaction energies [Table 4.8]. At the same time however, visual inspection and clustering analyses of the simulations indicated that the removal of the bulky phenyl ring allowed for increased ligand aglycone flexibility (heptyl α-D-mannoside) or rotation (biphenyl α-D-mannoside) [Figure 4.8]. Furthermore, given that the Tyr48 side chain was, in the case of wild type binding, forced to adapt a single, mode-dependent rotamer, a corresponding entropic cost was also to be expected. Consequently, the interplay of all these factors has likely resulted in a near-perfect entropy-enthalpy compensation for the Y48A mutation – a fact also apparent form the ITC data. As such, the Tyr48 residue should still be treated with utmost attention given the interactions it participates in for the wild type FimH.

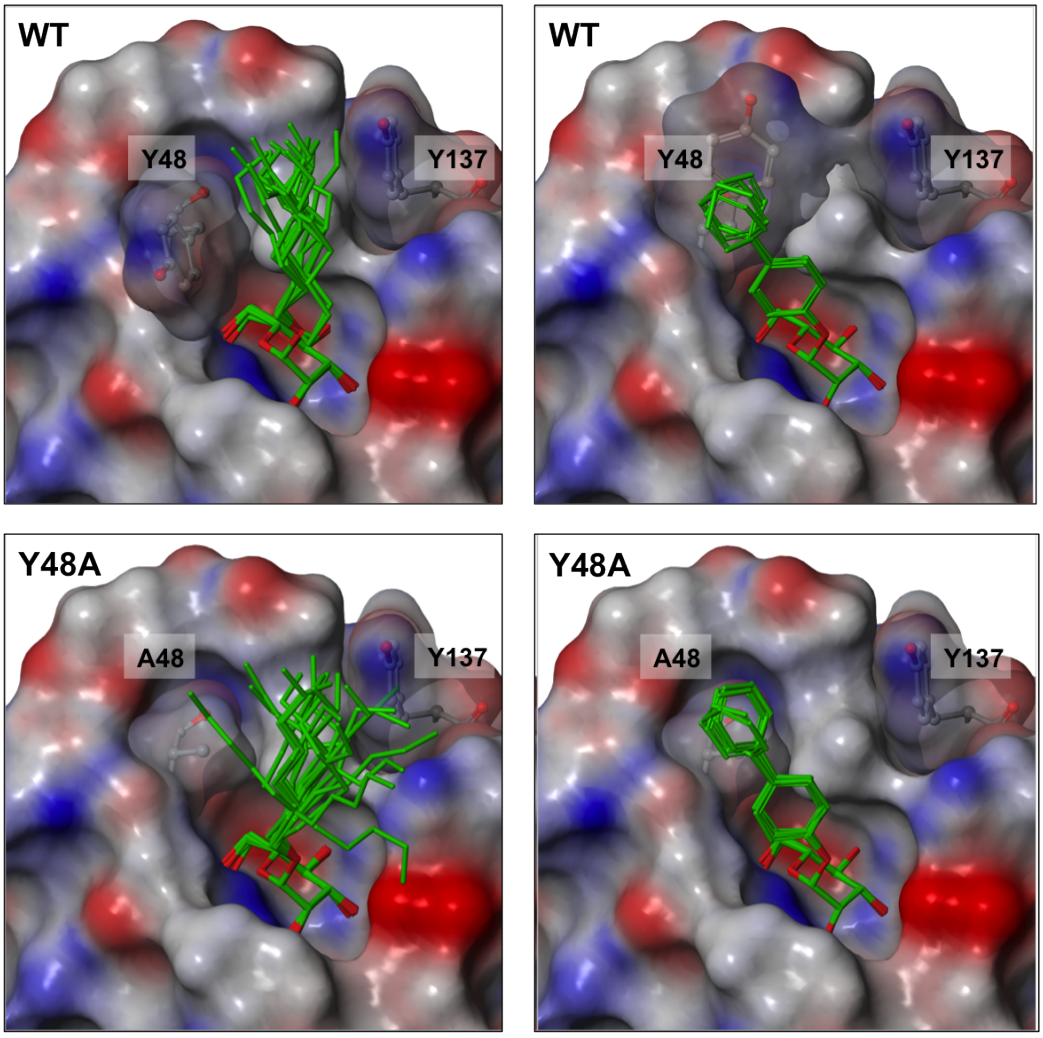


Figure 4.8: MD snapshots showing the binding of heptyl (left) and biphenyl (right) α-D-mannosides to the FimH LD wild type and the Y48A mutant. In both cases, the mutation resulted in increased ligand freedom (i.e. flexibility of the heptyl moiety and rotation of the outer biphenyl ring). Surface coloring corresponds to hydrophobic (gray), positively-charged (blue), and negatively-charged (red) regions.

With respect to the Y137A mutation, no satisfactory explanation could be obtained from MM/GBSA post-processing of holo MD simulations. Furthermore, employing the state-of-the art Free Energy Perturbation (FEP) methodology also only served to confirm the lesser importance of the 137 residue, especially for the biphenyl ligand ($\Delta\Delta G_{Y137A}$ of 0.9 and 0.6 kcal/mol for heptyl and biphenyl derivatives respectively), leading to assume that the necessary answers could not be derived form the protein-ligand complexes.

Hence, additional MD simulations for the apo wild type and mutant proteins, coupled with a detailed geometric and energetic analysis were performed (work done by Dr. Sameh Eid). As a result, a significant deformation of the apo binding site (side chains only) of the mutant was revealed, as demonstrated by the evident distortion of residues Tyr48, Gln133, and Asp135 [Figure 4.9]. Thus, upon ligand approach, the original binding site conformation needed to be re-assumed, with certain residues (e.g. Asp135) not capable of fully reorienting. As a consequence, an additional cost for a protein induced-fit had to be paid for the Y137A mutation, as reflected by the experimental data.

To confirm our findings, we were provided with crystal structures of heptyl α-D-mannoside in complex with wild type, Y48A, and Y137A FimH (structures solved by Dr. Julie Bouckaert). The three complexes supported our expectations regarding protein stability and a retained carbohydrate-binding pattern, showing identical orientations of the mannose moieties. The heptyl aglycones were also found similarly oriented between respective structures, though for the Y48A mutation four distinct conformers could be observed. As with the modeling data, this indicated a pronounced increase of aglycone mobility with respect to the wild type binding, leading to a more beneficial ligand conformational entropy. Lastly, the Tyr48 side chains were found perpendicular to Tyr137, opposing the results of molecular modeling. As already discussed in section 4.1.1, this was likely caused by crystal packing effects and unrelated to ligand binding. Short MD simulations for each structure, allowed the Tyr48 side chain to reorient, enclosing the ligand aglycone between the *tyrosine gate* residues.

During our study, additional FimH protein mutants were also investigated. Firstly, a set of serine mutants (Y48S and Y137S) was expressed and tested but did not reveal anything conflicting with the alanine-mutant based conclusions. Given the different properties of the respective hydroxyl and methyl groups, this confirmed that the side-chains of the mutated amino acids no longer had any impact on the binding. Secondly, an R98A mutation was also explored, but yielded affinities virtually no-different from the wild type FimH. While of no relation to the *tyrosine gate*, this confirmed that the previously disputed (Section 4.1.1) interaction between Arg98 and the biphenyl aglycone, was indeed absent under physiological conditions.

In summary, the performed study confirmed the vital importance of the FimH *tyrosine gate* toward binding – in multiple modes – of mannosidic ligands. It also provided direct evidence of the previously-assumed significant entropy differences between various compounds. A more detailed insight could not be gained however, due to mutation-induced distortions of thermodynamical and conformational equilibria within the native binding interface.

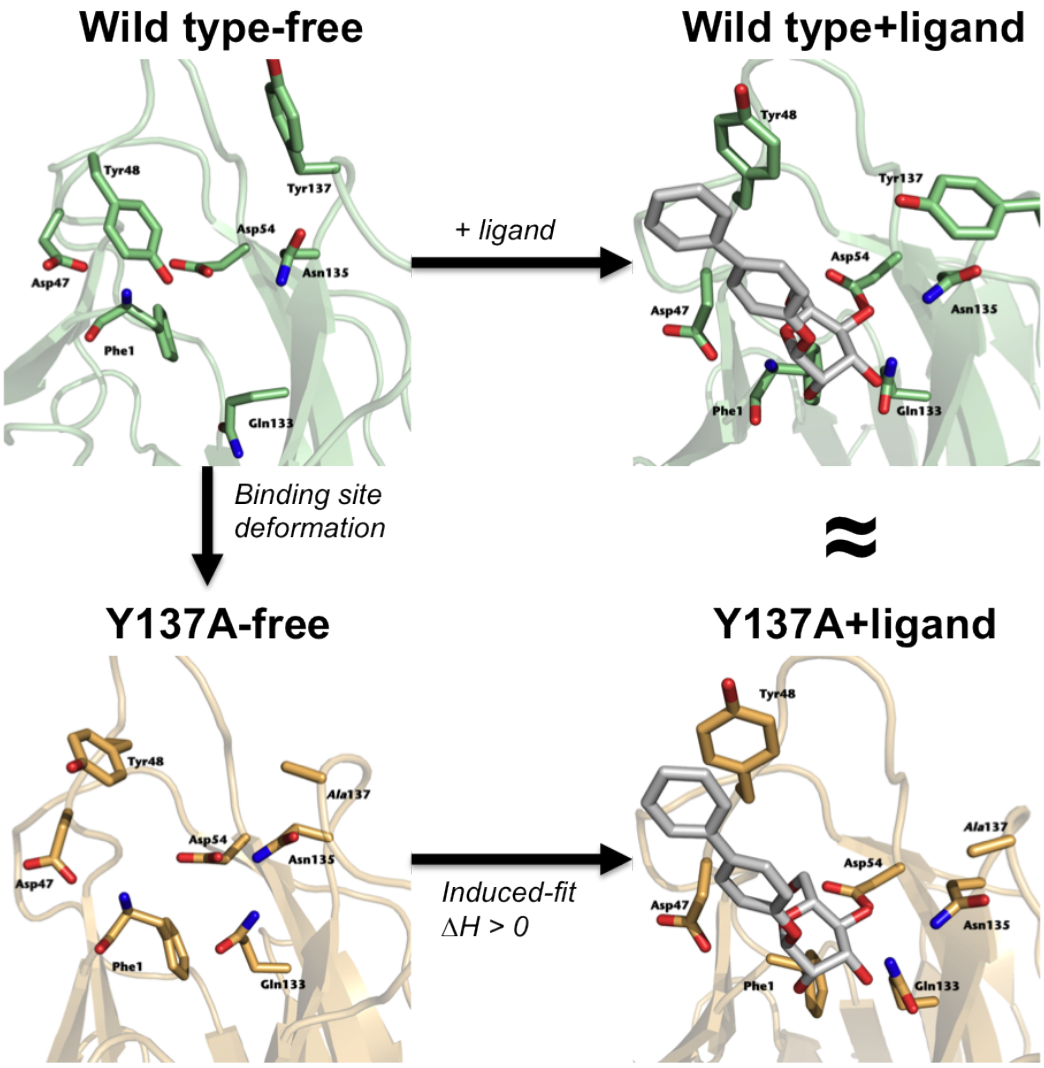


Figure 4.9: Snapshots from MD simulations of apo and holo FimH LD bound to the biphenyl α-D-mannoside. The bottom-left image shows the mutation-induced deformation of the wild type binding site (top-left). Upon ligand approach (horizontal arrows), the Y137A pocket needs to be rearranged (through an induced-fit) to accommodate the mannoside. The resulting interface conformation (bottom-right) still retains some difference from the wild type holo arrangement (top-right).

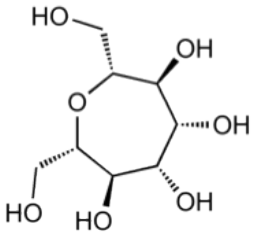
4.1.4 Mannose modifications

The substantial effort devoted into optimizing the aglycone portion of α-D-mannosides, has met with multiple limitations related to its maximum size, entropy-enthalpy compensation phenomena, and scarceness of interaction partners located outside of the mannose-binding pocket. At the same time, previous attempts of substituting or modifying the mannose moiety (including amino- and fluoro-substituted compounds synthesized and tested in our group),[160] resulted in substantially diminished affinities – a testament to the very high specificity of the FimH binding site. Nevertheless (as highlighted in Section 4.1.1), with roughly 60-70% of the binding enthalpy related to interactions with the carbohydrate moiety (likely over-compensating the entropy loss of the hydroxyl groups), modifying it, while challenging, held too much promise to be given up. In a renewed attempt to tackle the mannose specificity of FimH, additional compounds [Appendix Table 7.2] were designed and screened using the methods described in Section 4.1.1. Some of these attempts are discussed below.

Expanding the mannose ring

Given the failed attempts in substituting the mannose with other carbohydrates,[161] a conclusion was drawn that this moiety should be modified rather than replaced. In a first attempt, a heptose-like cyclic derivative was evaluated [Appendix Table 7.2: 1]. During the initial modeling, the bulky (with respect to mannose) structure could be docked into the binding pocket, forming eight hydrogen bonds. A subsequent MD simulation (2 ns) showed the docking pose to be stable, excluding the possibility of a too-high internal strain of the modified ring system. Finally, an *in vitro* evaluation performed competitively against the heptyl α-D-mannoside, yielded affinities in the low-micromolar ranges, providing a qualitative agreement with a modest but still clearly-favorable MM/GBSA score [Table 4.9]. Thus, while inferior to the previously discussed compounds, this derivative offered (to our best knowledge) the highest non-mannosidic ligand potency obtained at the time. As such, it served as proof of concept that the carbohydrate moiety could be, to an extent, modified. An additional observation gained from the modeling study, was a somewhat different orientation of the anomeric oxygen of the derivatives [Figure 4.10] that, through the addition of aglycones, could be exploited for further compound optimization.

Table 4.9: Binding characteristics of a derivative with an expanded carbohydrate ring, including constituents of the binding free energy (top; scaling coefficient 1/7), per-residue free energy contributions (right), and most stable hydrogen bonds (bottom). All results obtained either using *Amber* (averages over MD trajectories) or *Cheetah* and *Aquarius* software (docking; values in brackets) and given in kcal/mol and Å.



Energy	**Contribution**	**Residue**	**Contribution**
van der Waals	−18.1 (−12.0)	Phe1	−4.6
electrostatic	−70.9 (−28.7)	Ile13	−1.4
solvation	+41.0 (+13.5)	Cys44	−0.1
H-bond	n/a (−17.8)	His45	−0.3
total	−48.1 (−45.0)	Asn46	−0.8
		Asp47	−2.3
Exp. affinity	low-μM	Tyr48	−1.2
		Thr51	−0.1
		Ile52	−1.5
		Thr53	0.0

Acceptor atom	**Donor atom**	**% frames**	**Distance**	**Angle**	Asp54	−7.9
Ligand: O5	Phe1: N	70	2.9	156	Arg98	−0.1
Ligand: O5	Asp47: N	50	2.9	163	Gln133	0.0
Ligand: O3	Phe1: N	40	2.9	150	Asn135	−0.9
Ligand: O4	Asn135: ND2	30	2.9	164	Tyr137	−0.1
Ligand: O3	Phe1: N	10	2.9	152	Asn138	0.0
Asp54: OD1	Ligand: O4	100	2.6	165	Asp140	0.0
Asp54: OD2	Ligand: O5	100	2.6	169	Phe142	−0.5
Asp140: OD2	Ligand: O2	30	2.7	165	Phe144	−0.2

On a separate note, prolonged MD simulations performed for various FimH-ligand complexes indicated – as opposed to the image presented by the static crystal structures and docking – occasional reorientation of certain binding-pocket elements:

– Residue Asp140, interacting with the 3-OH group of the bound mannose in all crystal structures

– The so-called "clamp loop",[162] containing residue Ile13 positioned close to the C-2 of the carbohydrate moiety

While the first observation could be attributed to insufficient stabilization of the underlying force field and was circumvented by employing different software, the loop motion persisted under various setups under room temperature. The finding became even more intriguing in context of a compressed

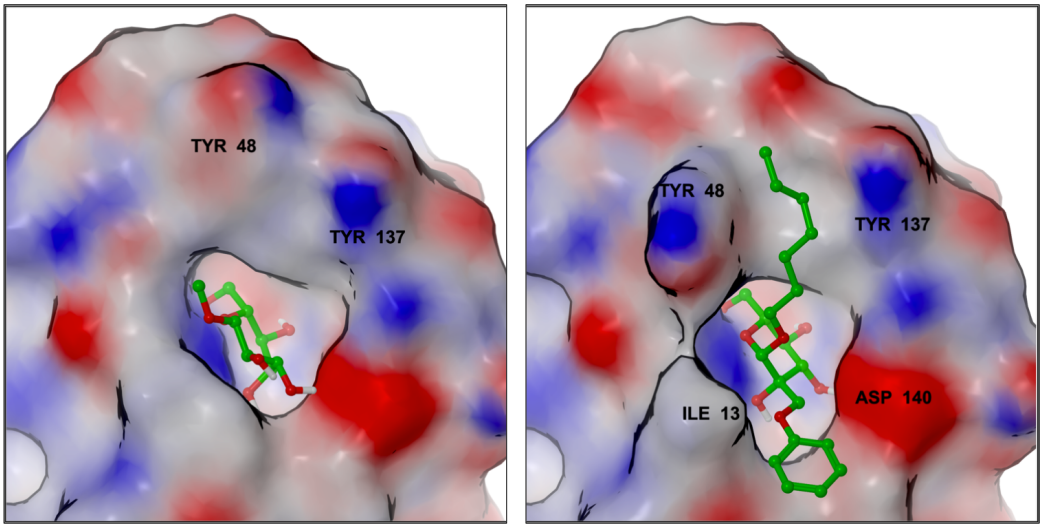
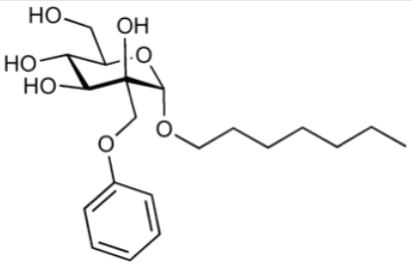


Figure 4.10: MD-refined docking poses obtained for the heptose-like (left) and a C-2 substituted (right) mannosidic derivatives. In both cases all crucial components of the hydrogen bond network could successfully be retained.Surface coloring corresponds to hydrophobic (gray), positively-charged (blue), and negatively-charged (red) regions.

FimH crystal structure (discussed in section 4.3), also showing a different position of the "clamp loop", leaving the C-2 mannose vicinity practically devoid of interactions. Spurred by this, an effort was made to expand the mannose in the direction of the potentially absent/flexible loop. For them to be successful, the new derivatives had to maintain an intact hydrogen-bond network. Thus, a series of derivatives with equatorial C-2 attachments preserving all initial hydroxyl groups, was docked and scored *in silico* [Table 4.10; Appendix Table 7.2: 2–6].

While this sub-project is still in an early stage (personal communication – Wojciech Schönemann), initial modeling indicated that the modified compounds could indeed maintain all discussed mannose and aglycone interactions while orienting the attached groups toward a previously not targeted region of the protein [Figure 4.10]. Subsequent evaluation by means of the competitive polymer binding assay, yielded potencies in the mid- to low-nanomolar ranges – best till date for non-mannosidic binders.[163] Thus, this attempt presents yet another successful modification of the mannose moiety, that could furthermore enable inhibition of both known conformational states of FimH.

Table 4.10: Binding characteristics of a derivative with an equatorial C-2 addition, including constituents of the binding free energy (top; scaling coefficient 1/7), per-residue free energy contributions (right), and most stable hydrogen bonds (bottom). All results obtained either using *Amber* (averages over MD trajectories) or *Cheetah* and *Aquarius* software (docking; values in brackets) and given in kcal/mol and Å.



Energy	**Contribution**	**Residue**	**Contribution**
van der Waals	−29.4 (−19.5)	Phe1	−3.4
electrostatic	−77.3 (−39.1)	Ile13	−2.8
solvation	+44.4 (+29.3)	Cys44	−0.1
H-bond	n/a (−16.5)	His45	−0.4
total	−62.3 (−45.8)	Asn46	−1.0
		Asp47	−2.1
Exp. affinity	147.3 nM	Tyr48	−2.3
		Thr51	−0.3
		Ile52	−2.1
		Thr53	0.0

Acceptor atom	**Donor atom**	**% frames**	**Distance**	**Angle**	Asp54	−7.7
Ligand: O3	Phe1: N	90	2.9	154	Arg98	0.0
Ligand: OH	Phe1: N	70	2.9	154	Gln133	−1.5
Ligand: O1	Gln133: NE2	50	2.9	164	Asn135	−1.6
Ligand: O2	Asn135: ND2	50	2.9	161	Tyr137	−0.6
Ligand: O3	Asp47: N	30	2.9	163	Asn138	0.0
Asp54: OD1	Ligand: O2	100	2.6	168	Asp140	−0.5
Asp54: OD2	Ligand: O3	100	2.6	167	Phe142	−1.1
Asp140: OD2	Ligand: O1	60	2.7	163	Phe144	−0.2

Replacement of a conserved water molecule

The highly selective conformation of the FimH binding pocket (in its high-affinity state) impaired modification of the carbohydrate hydroxyl groups. A potential foothold toward overcoming these drawbacks related to a conserved water molecule present at the deep end of the pocket in multiple apo and holo crystal structures (16 out of 35 analyzed protein chains belonging to 8 separate PDB entries). The displacement of such a molecule would be expected to result in enthalpic and entropic (up to 2 kcal/mol) rewards alike,[164] while offering a structurally new antagonist species, with possibly altered drug-like properties. Hence, an attempt was made to design ligand extensions at the C-2 mannose position, capable of displacing the water while mimicking its interactions with the protein.

Before proceeding with compound design, the importance of a hydro-

gen bond between the water molecule and the mannose 2-OH group had to be determined [Figure 4.11]. To this end, molecular dynamics simulations performed on various previously-discussed systems were "re-visited". A resulting statistical analysis of the distances between the donor and acceptor oxygens, yielded a mean value of 3.1 Å as opposed to 2.7–2.8 Å distances found through a search in the *Relibase* crystal structure database.[165] Additionally, simulations of the apo protein showed a frequent exchange of the water molecule with the bulk solvent. Hence, as long as all other mannose interactions would be retained, this bond – if at all present – was deemed an acceptable loss. As a result, a compound with the 2-OH group replaced with a 2-hydroxyethoxy moiety was selected for synthesis (personal communication – Dr. Katharina Mayer) and evaluation.

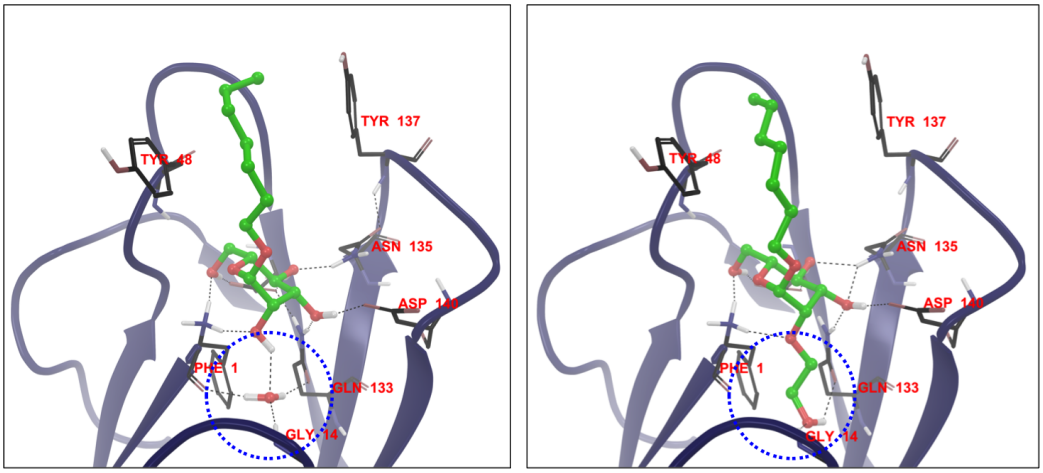


Figure 4.11: Top-scored docking poses for heptyl α-D-mannoside (left) and its derivative extended with the hydroxyethoxy moiety (right). Blue circles indicate the regions occupied by the conserved water molecule/ligand extension.

Flexible docking and subsequent MD simulations for the new ligand – performed in absence of the water justified by the above-mentioned observations –, confirmed it capable of maintaining a stable pose that could mimic most of the solvent molecule's interactions [Figure 4.11]. Binding free energy calculations, coupled with an RRHO entropy estimate [Table 4.11; Appendix Table 7.2: 7] showed equal promise. In fact, the calculated energy improvement toward heptyl mannoside was large enough ($\Delta\Delta G$ of 12.8 kcal/mol) to almost insure superiority of the modified compound independently from method inaccuracies (incomplete entropy and desolvation treatment discussed previously). Still, a subsequent evaluation by means of the competitive polymer binding assay and ITC, indicated an affinity several hundred times worse

than that of heptyl mannoside (difficult to determine more precisely given the assay discrepancies and detection limits).

Table 4.11: Ligand properties and MM/GBSA results obtained from respective simulations of heptyl α-D-mannoside and its extended derivative (in its most favorable pose). As opposed to other MD-based results, these were obtained using the *Desmond* (simulation), *Prime* (MM/GBSA), and *MacroModel* (RRHO entropy) software, enabling more insight into the energetic components of binding. All values in kcal/mol and Å^3. Energy components rounded to the first decimal digit.

	Heptyl α-D-mannoside	**Modified derivative**
Ligand volume	264	306
Number of heavy atoms	19	22
Hydrogen-bond donors	4	4
Hydrogen-bond acceptors	6	7
MM/GBSA $\Delta G_{Coulomb}$	-31.9	-37.5
MM/GBSA $\Delta G_{Lipophilic}$	-30.6	-37.8
MM/GBSA $\Delta G_{Solvation}$	18.8	19.0
MM/GBSA ΔG_{vdW}	-23.7	-28.9
MM/GBSA ΔG_{total}	-67.1	-81.8
Ligand strain	4.5	6.5
Protein strain	11.1	12.4
RRHO entropy	17.8	19.7
ΔG_{total}	**-49.3**	**-62.1**

Given that the obtained *in silico* data strongly indicated that the ligand – in its favorable conformation – should benefit from a substantial increase in binding energy, we concluded that it failed in displacing the water molecule. In order to account for this scenario, docking to a protein grid retaining the molecule was performed. As a result, a pose with an inferior score and a noticeable distortion of the hydrogen-bond network was obtained [Figure 4.12]. A subsequent MD simulation resulted in further pose deterioration, with more crucial interactions with the protein being lost (e.g. with residue Phe1). Interactive inspection of the simulation revealed that the buried water molecule remained immobilized by the added group and impaired optimal binding. The origin of the unfavorable initial geometry could be discovered by means of ligand-only molecular dynamics, in which conformations close to the one required for favorable binding were scarcely populated [Figure 4.12].

Based on the performed study, we concluded that the modified ligand could neither approach the receptor in a favorable conformation nor adapt a water-displacing shape within the binding site due to steric clashes. It remains possible that more pre-organized carbohydrate extensions could prove more successful, all the more so since the results reported in Section 4.1.4 confirmed the relatively limited significance of the mannose C-2 vicinity. A crucial factor to be kept in mind however, is the once again encountered significance of entropy toward the overall ligand binding. At the present the synthesized compound is also pending evaluation using the full, two-domain FimH protein that could yield a different result.

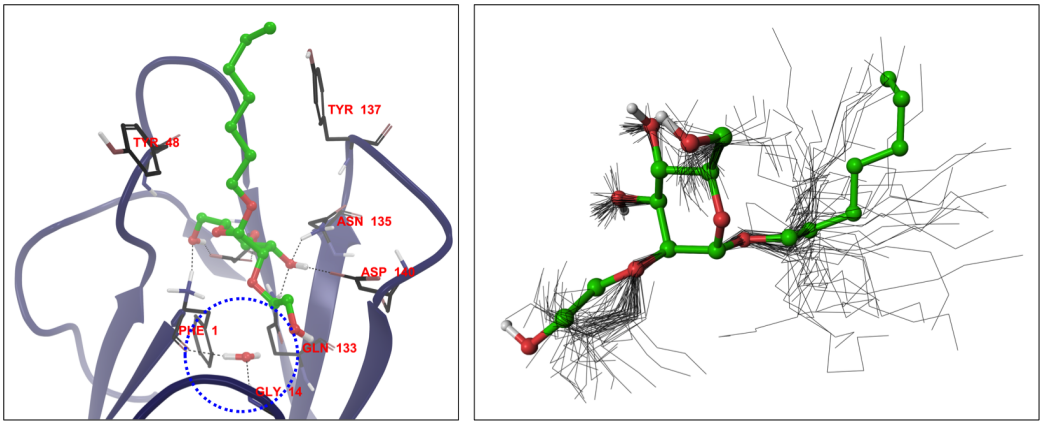


Figure 4.12: (left) Binding pose of the hydroxyethoxy-extended derivative, assumed upon failing to displace the conserved water molecule (position indicated by the blue circle). (right) Free state conformer distribution of the extended compound (black lines) in comparison with the conformation necessary to displace and mimic the binding site solvent (ball and stick).

4.2 The *VirtualDesignLab*

During the course of modeling the various compounds presented in the previous sections, several limitations of the employed computational methods became apparent with respect to their precision. Firstly, given the discovery of multiple FimH binding modes, selection of single structures representing each ligand could have led to errors. Furthermore, the implicit solvation treatment implemented in the *Induced Fit Docking* and MM/GBSA protocols, could have proven inappropriate for simulating the highly solvent-exposed FimH binding site. Finally, given the inherent flexibility of the protein-ligand interface, the employed treatment was disregarding the vital role of the binding entropy. It was also concluded that the tedious and repetitive procedures involved in the evaluation of various compound series were too inefficient and prone to human error. In order to circumvent all of these limitations, an automated framework – the *VirtualDesignLab* – was developed, consisting of design, docking, and scoring modules derived from technologies of the *VirtualToxLab* platform. Due to its original purpose as well as consistency of this work, the framework will be presented as applied toward design of FimH ligands, focusing on training/test compound selection and validation of the resulting QSAR model (responsible for affinity predictions), along with later attempts of more accurate entropy calculation. Still, it bears mentioning that the underlying technology enables re-orientation toward other targets as discussed at the end of this section. Also, a more general description of the framework is presented as a part of a recently published article.[139]

4.2.1 Docking module and base compound selection

Although the state-of-the-art *Induced Fit Docking* protocol has proven sufficient for screening purposes, it still showed limitations toward handling the structurally-diverse FimH inhibitors. Hence, a decision to replace it in favor of the *Alignator* and *Cheetah* software (performing pre-alignment and docking tasks respectively) was reached, due the following benefits:

- Explicit solvent treatment, enabling more precise assessment of related properties and incorporation of structural water molecules
- Employment of a directional force field, improving the treatment of hydrogen bonds
- A feature of using multiple templates to guide the flexible alignment and docking toward specific, crystal-structure based binding modes

– Both programs being integral and thoroughly validated components of the parent *VirtualToxLab* technology

Before employment, the new protocols were evaluated with respect to handling mannose-based ligands. A secondary objective here was to confirm the necessity of later employing the mQSAR-based scoring and – if so – obtaining test and training substance sets for the development of the FimH model. To both ends, a set of 108 FimH inhibitors with known experimental affinities was docked [Appendix Table 7.3]. These compounds were carefully selected in order to represent maximum diversity in terms of structure and potency (spanning nearly five orders of magnitude from 220 μM to 2.4 nM) and contained representatives of all structural classes presented in Sections 4.1.3 and 4.1.4. The generated ligand geometries confirmed that the newly employed software was capable of correctly docking the mannose moieties and reproducing the surrounding hydrogen-bond network. Sampling of multiple Tyr48 conformations could also be observed, indicating that different binding modes were being accounted for. Nevertheless, pair-wise comparison with the top-scored poses generated by *Induced Fit Docking* protocol yielded RMSD values between 0.2 Å and 6.1 Å, with a mean of 2.0 Å. This clearly indicated that the two methods were at times favoring significantly different conformations, none of which could be discarded given the pronounced aglycone motions discovered previously. Also, between the singular docking scores and the experimental affinities, coefficients of determination (r^2) of 0.202 and 0.232 could be obtained for *Induced Fit Docking* and *Alignator/Cheetah* respectively. Given that docking protocols are commonly known to generate reliable geometries while suffering from inaccurate scoring/ranking,[166] this confirmed the necessity of using multiple ligand poses along with a more accurate affinity estimation method. As such, up to eight geometries per-ligand were used for the FimH mQSAR model development and all later predictions.

4.2.2 QSAR model

The FimH QSAR model currently employed for the *VirtualDesignLab* was selected from a series established based on structural and biological data of the above-mentioned 108 mannosidic inhibitors displaying diverse PK/PD profiles. Compound synthesis, biological assays, and model development were performed in-house, ensuring consistency of all results. Table 7.3 of the Appendix shows the structures and experimental/predicted affinities of compounds used for the development. The model based on a genetic algorithm implemented in the *Quasar* software,[167,168] converged at a cross-validated r^2 of 0.805 and yielded a predictive r^2 of 0.596 [Figure 4.13]. The only modest

value of the predictive r^2 was a consequence of the relatively narrow range of test compound affinities, with some substances necessary for the training set due to their structural uniqueness. The performance of the model was therefore better reflected by the individual predictions, with 23 out of 24 test substances found within a factor of 10 from their experimental affinity. Thus, it could be stated that the method displayed predictive power vastly superior to the previous attempts. The FimH model was then further challenged by using Y-scrambling and consensus scoring with the software *Raptor* (dual-shell 5D-QSAR; Figure 4.13).[138] All tests, including the processing of additional compounds, further confirmed the predictive power of the mQSAR model and the framework established around it.

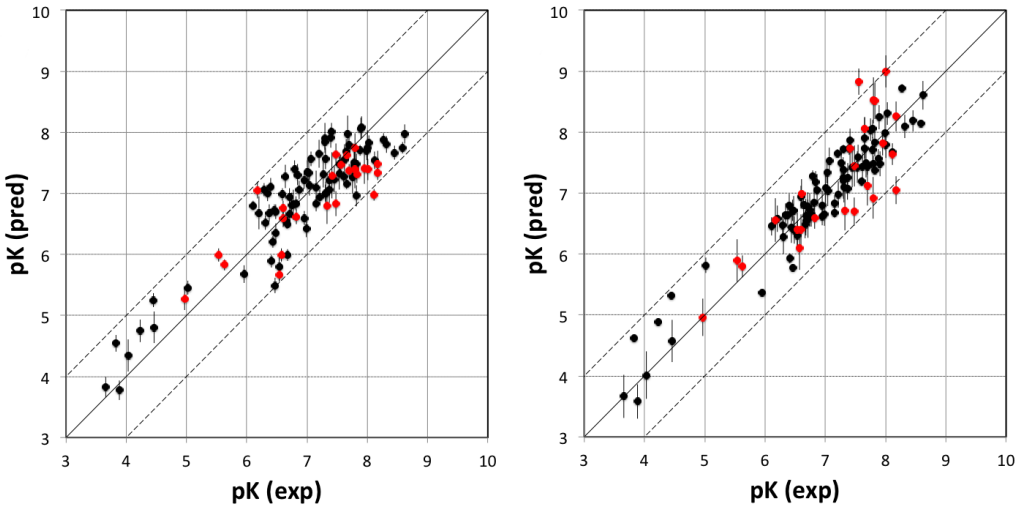


Figure 4.13: Comparison between experimental (horizontal axis) and predicted pK values (vertical axis) for the *Quasar* (left) and *Raptor* model (right). Black and red points represent compounds of the training and test set, respectively. Vertical bars indicate the estimated standard deviations of the prediction. Dashed lines are drawn at factors of 10 from the experimental value. The difference between the two methods relates to the latter lacking terms for electrostatics and polarization (replaced with a hydrophobicity term) and incorporating a dual-shell representation of the induced-fit.

4.2.3 Entropy treatment

The entropy change related to various association events is an often important and at times crucial – due to serving as the driving force – component of the binding energy. Nevertheless, experimental means of measuring these

changes are still limited to complex and costly ITC and NMR techniques,[169] impairing many drug design and optimization attempts. This holds equally true for *in silico* approaches that – despite decades of research – are still primarily centered around extremely computationally expensive Thermodynamic Integration and Normal Mode Analysis [NMA] techniques.[170] Thus, within the scope of higher throughput evaluation, simple attempts of approximating ligand binding entropies are employed, commonly based on a constant penalty per a rotatable bond (around 0.5–0.7 kcal/mol). While the addition of this term has at times proven beneficial,[171] this was likely not to hold true with respect to modeling of prospective FimH ligands. Specifically, the previously-discussed inherent flexibility of the protein-ligand interface was likely to be treated with a too high penalty that assumed full rigidification of all rotatable bonds. Due to such, a method for individually scaling the per-bond penalties – based on probing the rotation accessible to each bond in the bound and free ligand states – was developed for the purposes of the *VirtualDesignLab*. The conceived technology additionally enabled – by employing the RRHO approximation – calculating the entropy components related to the loss of rigid body translation, rotation, and vibration, commonly neglected in computational and experimental approaches alike.

Following development and testing (see Section 3.9), the method was employed toward establishing a new QSAR model for the FimH adhesin. The resulting receptor surrogate – developed otherwise identically to the one presented in the previous section – converged at a cross-validated r^2 of 0.830 yielding, a predictive r^2 of 0.640 [Figure 4.14], with all test set compounds within a factor of 10.0 from their experimental affinities. Given the identical settings and set compositions (along with their discussed limitations), this was considered a clear improvement over all previously tested models. In order to exclude the possibility of the calculated values coincidentally benefiting the *Quasar* technology, the new software was also coupled with the MM/GBSA protocol, by means of multiple linear regression (necessary given the heterogeneous origin of the data). To this end, a separate set of compounds was used [Table 4.12], offering a good correlation with experimental ITC data (r^2 of 0.804) and a large variation of binding entropy changes ($T\Delta S$ values spanning a range 3.5 kcal/mol). As a result, a coefficient of determination of 0.846 could be obtained, with the entropy results differing by up to 3.8 kcal/mol from one another. All in all, a satisfactory performance of the developed method could be obtained along with indications toward its general applicability outside the *VirtualDesignLab* framework. With the software also employed toward establishing a carbohydrate scoring function (personal communication – Dr. Sameh Eid), this applicability can hopefully soon be shown in a dedicated publication.

On a side note, it should be mentioned that when using only parts of the calculated entropies (conformational, vibrational, rotational, or translational), the prediction quality rarely surpassed the constant/no penalty approaches – a likely testament toward the highly correlated nature of the state function's components. This led to think, that accounting for the third primary entropy constituent – solvation – could further improve the quality of the results. Regrettably though, the complex nature of the calculations related the explicit solvent along with the necessity of its dynamic representation, impaired developing a suitable method within the scope of this work. As it stands, the conceived technology is best employed toward series of congeneric ligands such as the FimH binders, where the bulk of solvation changes concerns the fully desolvated, identically bound mannose.

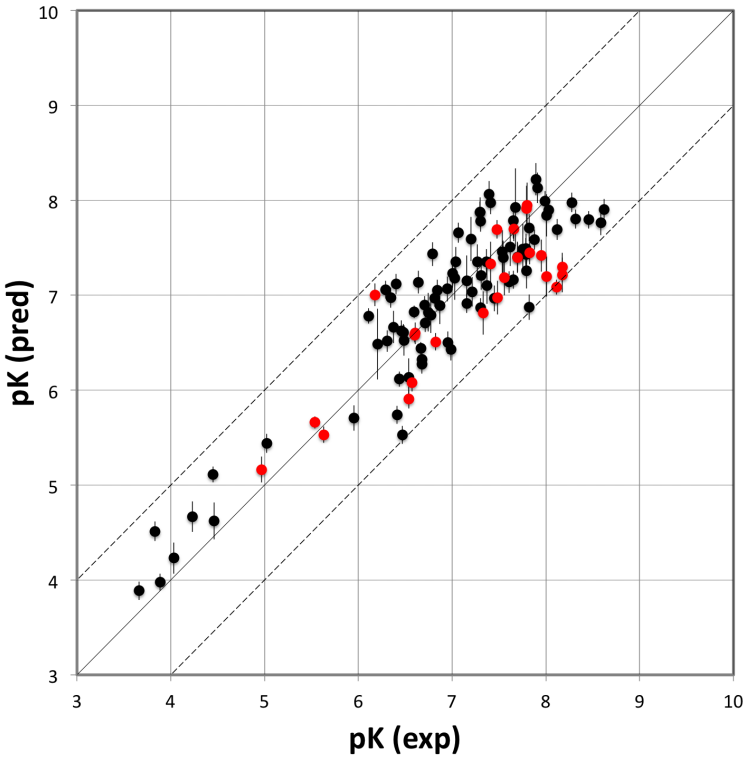
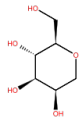
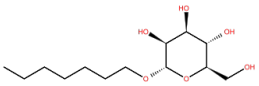
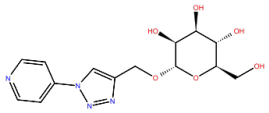
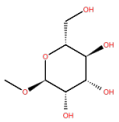
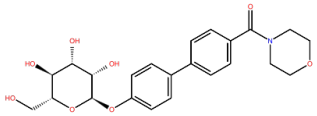
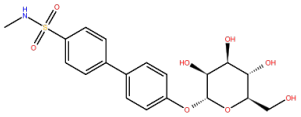
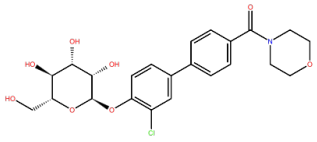
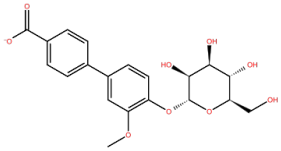
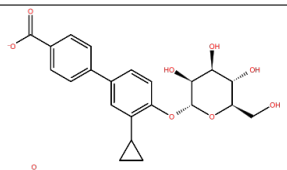
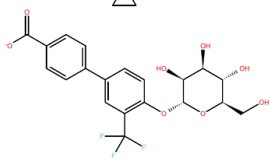
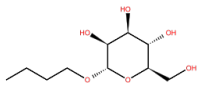
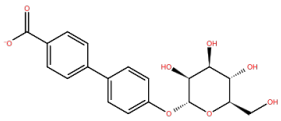
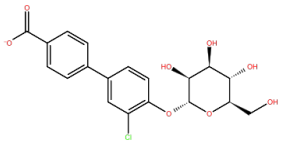


Figure 4.14: Comparison between experimental (horizontal axis) and predicted pK values (vertical axis) for the *Quasar* model employing the improved entropy estimation. Black and red points represent compounds of the training and test set, respectively. Vertical bars indicate the estimated standard deviations of the prediction. Dashed lines are drawn at factors of 10 from the experimental value.

Table 4.12: Experimental and calculated free energies (MM-GBSA; software *Prime*; scaling coefficient 1/10) for a small set of ITC-tested compounds, with and without the calculated entropy estimate. All values in kcal/mol and rounded to the first decimal digit. Entropies given as –TΔS values for 298 K.

	Structure	Exp. free energy	Predicted free energy	Trans./rot. entropy	Conform. entropy	Corrected prediction
1		−8.1	−68.6	16.7	2.5	−49.4
2		−10.3	−98.8	19.7	4.6	−74.5
3		−9.1	−90.2	18.4	3.9	−67.9
4		−8.9	−71.2	16.4	2.7	−52.1
5		−11.4	−100.6	19.6	3.7	−77.3
6		−11.6	−102.3	20.1	3.5	−78.7
7		−11.9	−106.8	20.3	3.3	−83.2
8		−10.9	−92.8	18.3	3.0	−71.5

Continued from previous page ...

	Structure	Exp. free energy	Predicted free energy	Trans./rot. entropy	Conform. entropy	Corrected prediction
9		−11.1	−100.3	18.2	5.0	−77.1
10		−11.5	−94.0	18.2	3.0	−72.8
11		−9.4	−82.2	17.3	3.1	−61.8
12		−10.7	−89.7	17.9	2.9	−68.9
13		−11.4	−98.9	18.3	4.7	−75.9

4.2.4 Performance, limitations, and re-purposing

The *VirtualDesignLab* is aimed at predicting the binding affinity for a given compound within a factor of 10.0 from the experimental value. Currently, the prediction for a single candidate requires approximately one hour of CPU time (with additional 3–5 minutes needed when employing the improved entropy estimation), offering a good balance between accuracy and processing time. Nevertheless, as can be seen in the Appendix Table 7.3, the precision of the underlying models tends to weaken for highly flexible structures. In such cases, a non-static approach offered by MD simulations is often still helpful.

It should also be emphasized, that the framework, while initially developed for the purposes of FimH inhibitor design and evaluation, is independent of the FimH mQSAR model. As such – assuming that a satisfactory amount

of ligands with experimental affinity data is available –, it only requires the generation and validation of a new model to enable re-orientation toward any target of interest. Furthermore, as long as an experimental structure or a homology model of the target exist, the corresponding docking and scoring tasks may be accomplished through numerous, freely accessible computational tools, making the framework intendant form any particular piece of software. Still with respect to the QSAR model development and validation, the procedures are best handled using the originally employed *Quasar* software to ensure satisfactory performance.

4.3 Modeling on the complete FimH protein

As highlighted in Section 2.4.3, the majority of research performed on FimH was, until recently, based on its detached lectin domain, rather than the complete structure. While mostly dictated by the difficulty of expressing the full protein in its native state (instability of the PD in absence of the FimG/FimC donor strand), this approach could be justified by a general correlation between *in vivo* and *in vitro* results (with the latter obtained using single-domain constructs).[145] Still, given the limited success in further improving the affinities of alkyl and biphenyl mannosidic derivatives, insight into the full structure of the receptor along with its dynamics (especially under no-flow conditions) remained imperative. In pursuit of this, several studies were performed during the last decade,[172] revealing that the protein could assume multiple conformations (resulting in varying ligand affinities) depending on several conditions:

- Presence of sheer force exerted by the physiological urine flow (affinity dependent on the sheer level, resulting in slip/catch-bonding characteristics)
- Insertion of the FimC chaperone between the respective domains (high affinity)
- Absence of the pilin domain (high affinity)
- None of the above (low affinity; refereed to as self-inhibition)
- Natural mutations (ligand affinity dependent on the specific mutation)

Based on these findings, the crystal structures available up until a few years ago – given the FimC-stabilization or lack of the PD – were associated with a high (most likely maximum) affinity FimH state, leaving the native apo protein conformation unknown. More recently though, it could be argued that a high-affinity conformation was also assumed under static conditions in presence of soluble mannose,[173] implying an induced-fit mechanism and supporting the validity of single-domain based work. Specifically, the study stated that the presence of the carbohydrate ligand enabled the binding of selective antibodies (expected to attach to an interface within the inter-domain region in a manner similar to the FimC chaperone) that would otherwise show no affinity toward the whole protein. Even more recently however, a crystal structure of the type-1 pili tip (with the common F18 variant of the FimH lectin) could be solved,[95] for the first time enabling analysis of the receptor

in its native apo state. As a consequence, a drastically different protein conformation could be observed with respect to ones previously obtained. In the context of the previous findings and assumptions, this discovery would have meant that the FimH protein could undergo an unprecedentedly-large conformational change under flow and static conditions alike. Given the up to 200-fold affinity variations reported for different protein variants and experimental conditions,[174] this made the determination of the underlying mechanism all the more crucial. Furthermore, it was still necessary to confirm whether one, both, or neither of the known structure types offered a valid physiological representation of FimH and whether this representation differed between flow and static conditions (the latter relevant for inhibition attempts due to the compounds not being attached to the bladder surface).

4.3.1 Structure preparation and analysis

In order to obtain representative structures of the two FimH states and enable their energetical comparison, constructs based on the 3JWN and 1KLF crystal structures (with 2.7 Å and 2.8 Å resolution respectively) were generated. Specifically, the 1KLF structure was modified to comply with the F18 amino acid sequence of the 3JWN counterpart (mutations V28A, S70N, and N78S). It was also stripped of its accompanying FimC chaperone (natively detached once the pilus rod assembly process is concluded), aside from a short strand of residues stabilizing the pilin domain. Similarly, the other conformer was separated from the remaining pili tip domains, with the exception of the FimG donor strand. Following preparation, the conformers were subjected to 10 ns MD simulations, given the previously noted possibility of crystal packing distortions and the necessity of relaxing the modified amino acid sequence of the 1KLF construct in absence of FimC.

While the performed simulations did not reveal significant changes within the domains of either state, a shortening of the distance between the 1KLF subunits could be seen. This was due to a "hydrophobic collapse" following the FimC removal, and gave reason to believe that this fully extended FimH conformation could not exist without shear flow or another stabilizing element. As such, structures representing the most stable parts of each MD trajectory were selected for further work [Figure 4.15]. Geometrical analyses of these representatives confirmed the retained structural differences initially observed with the crystal structures. Specifically, the 1KLF construct was found nearly 20 Å longer, with the LD and PD only weakly connected. The native apo structure on the other hand, presented a tighter contact between the subunits enabling multiple – primarily hydrophobic – interactions, as described in a publication released alongside the crystal structure.[95] RMSD

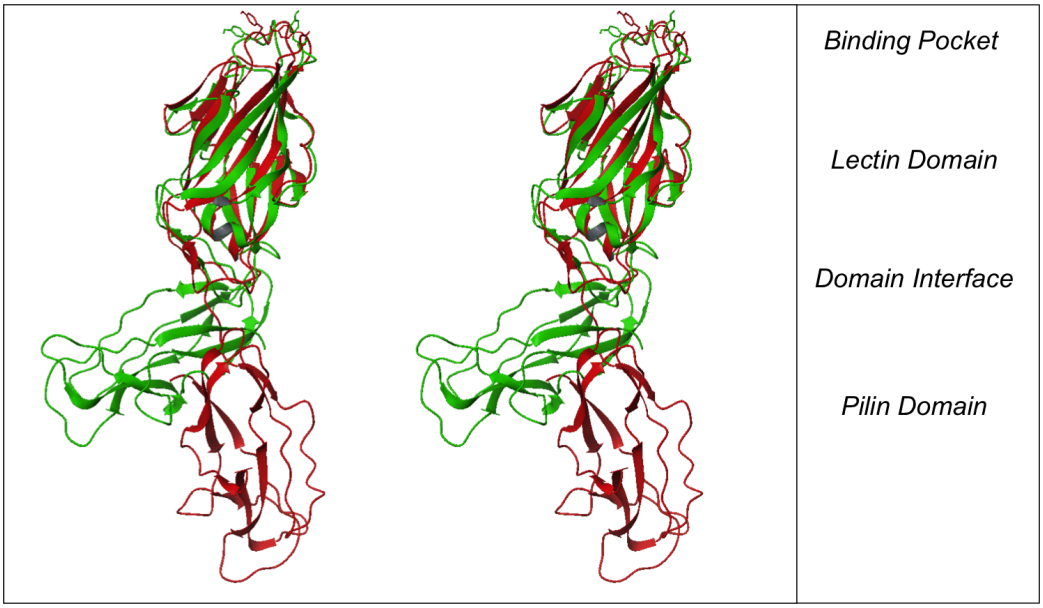


Figure 4.15: Stereo representation of MD-refined structures of the native apo (green; PDB code 3JWN) and FimC-stabilized (red; PDB code 1KLF) FimH protein structures, superimposed on the C_α atoms of the lectin domains. Upon closer inspection, one may observe that the dihedral angle between the domains differs by nearly 180 degrees.

calculations revealed that aside from the domain separation/compression, the most pronounced structural differences were focused in the LD (3.6 Å for 116 C_α atoms outside of the domain interface, with the 1KLF LD longer by approximately 10 Å), whereas the PD was much more conserved between the states (1.8 Å for all 121 C_α atoms). Focusing on the mannose-binding region – defined using the heavy atoms of residues Phe1, Ile13, Asp47, Ile52, Asp54, Gln133, Asn135, and Asp140 – yielded an RMSD of 3.5 Å, with the 3JWN residues partaking in forming a large, nearly-flat surface, hardly resembling the well-defined 1KLF mannose-binding pocket and unlikely to prove mono-mannose selective [Figure 4.16]. While potentially related to the lack of a ligand in the 3JWN structure, the overall scale of the differences substantially surpassed those commonly related to induced-fit events,[175] contradicting the theory that binding of mannosidic ligands imposed a conformational transition between the two crystal structure conformations. Also, no explanation could be found as to how rearrangements of this magnitude (commonly related to μs or ms time scales) could occur quickly enough to accompany sub-μs binding events.

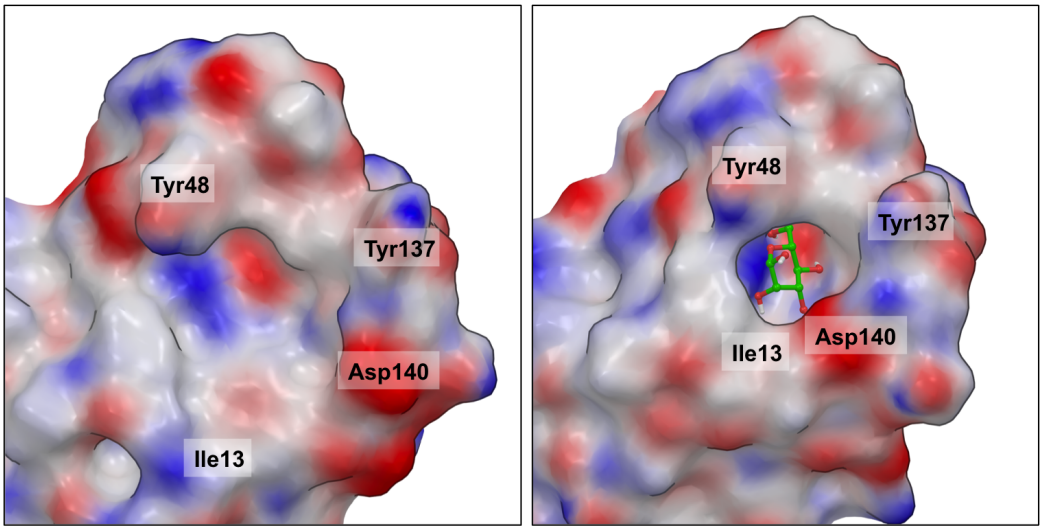


Figure 4.16: Binding site regions of the 3JWN (left), and 1KLF (right) crystal structures assumed to correspond to the low- and high-affinity FimH conformational states. Surface coloring corresponds to hydrophobic (gray), positively-charged (blue), and negatively-charged (red) regions.

An alternative explanation for the FimH binding characteristics, could be conceived in the form of a conformational selection mechanism. In light of this theory, large-scale changes of the target receptor were less unlikely as long as the relevant conformations could maintain an equilibrium. In order for the selection not to result in a kinetic bottleneck (with the concentration of the binding-capable species directly proportional to the k_{on} rates),[176] the underlying energy difference along with the size of the barrier between the states would have to be moderate however. To determine whether this scenario was more likely than the large-scale induced-fit mechanism, attempts were made to probe the behavior and energetics of relevant apo and holo conformations by means of MD simulations and MM/GBSA calculations employing the well-established AMBER-99SB force field (see Section 4.1).[177] Also, to avoid an incomplete free energy assessment, the calculations were coupled with the normal mode analysis (NMA) entropy estimation.

4.3.2 Protein energy calculations

The likelihood of preferring either the induced-fit or conformational selection mechanism by a protein may be derived from the underlying energy landscape formed by its relevant conformers. Specifically, surmounting a large energy barrier between the states requires a catalyst in the form of initial

ligand binding occurring as part of an induced-fit. A smaller barrier coupled with larger structural changes (rare for induced-fit due to the time necessary for them to occur) are commonly related to inherent protein motion enabling conformational selection. In light of this, probing whether the native "compressed" FimH form was capable of binding mannosidic ligands (without undergoing a large-scale change), was first required. If proven otherwise, a selection mechanism in which only one conformation could bind the ligands, would have became likely. To this end, the previously discussed *Induced Fit Docking* protocol – allowing residual motion necessary to escape the unfavorable planar binding site arrangement – was employed to dock a methyl α-D-mannoside (MM; chosen to focus conformational sampling of the docking protocol on the mannose binding) to the 3JWN-based grid. As a result, several complexes could be obtained, none of which reflected the binding pattern observed in the 1KLF structure. Furthermore, subsequent MD simulations would commonly lead to a complete ligand unbinding due to insufficient amount of favorable interactions with the protein (commonly 2–3 hydrogen bonds formed for the initial poses). Hence, assuming that the same residues were involved in mannoside binding to all FimH states, the ligand was docked manually to the 3JWN structure, with priority placed on reproducing as many 1KLF hydrogen bonds as possible. A resulting pose was then subjected to a short MD simulation while keeping the obtained interactions constrained. Finally, the constraints were removed and the system relaxed during an extensive, 100 ns simulation. As a result, the protein-ligand interface could be maintained, enabling reproduction of seven 1KLF hydrogen bonds (though lacking the charge-assisted bond with Asp140 as well as the hydrophobic interactions with the Ile13 "clamp loop"). Furthermore, a formation of a clearly defined pocket could be observed around the methyl mannoside [Figure 4.17], indicating that while the site was not natively preorganized, the compressed FimH state was still capable of ligand binding.

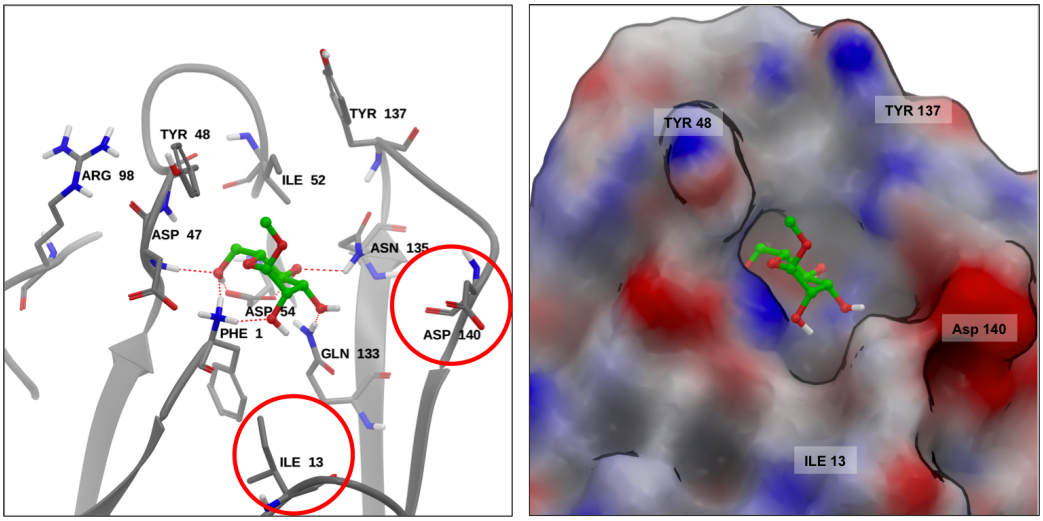


Figure 4.17: Cartoon (left) and surface (right) representations of the MD-refined MM binding pose with the native FimH protein. The resulting interface shares a lot of resemblance to its highest affinity counterpart but lacks – resulting in a somewhat larger binding site – a charge-assisted hydrogen bond with Asp140 and hydrophobic interactions with the "clamp loop" containing the residue Ile13. Surface coloring corresponds to hydrophobic (gray), positively-charged (blue), and negatively-charged (red) regions.

Obtaining the alternative binding pose, enabled probing the energetics of the apo and holo conformers with the MM/GBSA method [Table 4.13]. As an expected result, a more favorable ligand binding free energy could be obtained for the elongated system (–26.5 and –12.5 kcal/mol for 1KLF and 3JWN constructs respectively). Still, while likely prone to overestimation, the protocol also revealed an approximately 21 kcal/mol more favorable total free energy of the holo 3JWN complex over the 1KLF. Furthermore, summing up the energies of the free ligand and the native apo protein, yielded a 6 kcal/mol lower value to that of the 1KLF complex. Taken together this indicated that the formation of a complex with the extended protein was not a thermodynamically favorable process (when started from the compressed FimH) and required an external force – physiologically provided through shear stress – to counter a strong hydrophobic effect "collapsing" the domains. Additionally it could also be shown that the native apo protein had a significantly lower energy than its higher affinity counterpart ($\Delta\Delta G = -32$ kcal/mol), undermining the theory that these two states could maintain an equilibrium.

Table 4.13: MD-based results of MM/GBSA (software *Amber*) free energy measurements performed for holo and apo FimH systems. Values may be compared due to imposing a 100% sequence identity between the systems. Entropy estimates obtained from normal mode analyses (NMA) at 298 K. All values in kcal/mol and rounded to the first decimal digit.

	Compressed state	**Elongated state**	**Residual**
Receptor $_{MM/GBSA}$	-23000.5	-22958.1	-42.3
Receptor $_{NMA}$	2959.6	2969.8	-10.2
Receptor $_{total}$	-25960.0	-25927.9	-32.1
Ligand $_{MM/GBSA}$	-112.2	-115.3	3.0
Ligand $_{NMA}$	34.7	34.8	-0.1
Ligand $_{total}$	-146.9	-150.0	3.1
Complex $_{MM/GBSA}$	-23147.4	-23113.0	-34.4
Complex $_{NMA}$	2975.2	2988.3	-13.1
Complex $_{total}$	-26122.6	-26101.3	-21.3
ΔG $_{MM/GBSA}$	-31.6	-42.6	11.0
TΔS $_{NMA}$	-19.1	-16.1	-3.0
ΔG $_{total}$	-12.5	-26.5	14.0

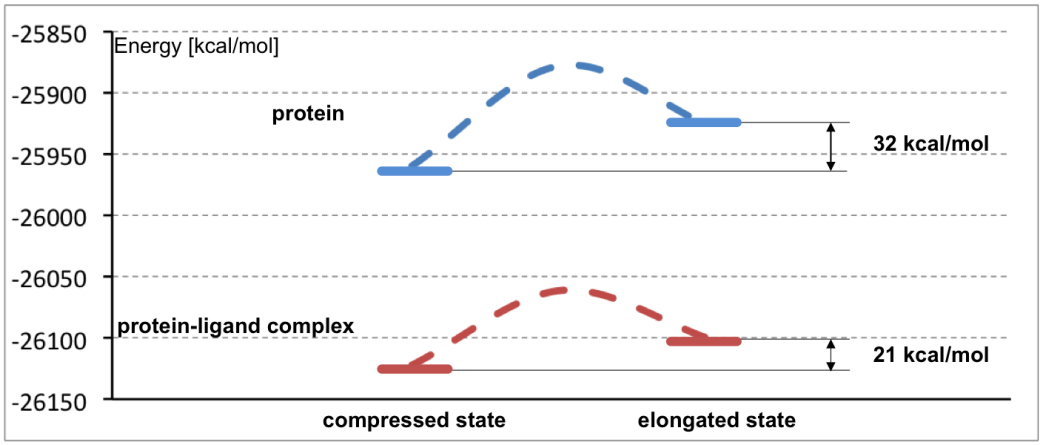


Figure 4.18: Graphical representation of the energy values obtained for compressed and elongated states of the FimH protein (blue) and the FimH-MM complex (red). Even though the ligand binding lowers the energy of both states and the difference between them, the native protein arrangement is favored in both cases.

4.3.3 Steered molecular dynamics

The obtained results strongly opposed the possibility of FimH transiting between its two crystal structure conformations without the influence of external forces. Nevertheless, it could not be excluded that the employed energy scoring methods were too inaccurate to compare the states. Furthermore, given lack of structural data regarding mannosidic ligand binding to the compressed protein form, the validity of the obtained docking pose could not be confirmed. Thus, a concept was devised to support the MM/GBSA results with steered molecular dynamics (SMD) simulations applying a pulling force on the system in a manner similar to atomic force microscopy experiments. Through this, the stability of the obtained ligand pose could be probed under flow conditions, reflecting the naturally occurring cell adhesion events in which the bacteria attach themselves to the highly-mannosylated cell surfaces. While of little relation to static inhibition attempts, an inability of simulating the accompanying domain separation or loss of ligand binding (preceding subunit detachment), would undermine the validity of the obtained pose and the derived conclusions. An additional benefit of the SMD simulations – if performed multiple times from closely-related structural starting points – was the possibility to plot the free energy changes along the pulling path (referred to as the potential of mean force or PMF). Through this, binding and barrier energies along with intermediate states involved in the transition could be highlighted, possibly delivering evidence

toward induced-fit/conformational selection preferences. This type of simulations was already previously employed toward studying the FimH protein.[178] The past work however, was based on the apo protein structure and limited to a single, constant-force run, without deriving the underlying energetics. Hence the study presented here remains – to my knowledge – unprecedented.

The performed simulations (run under a constant pulling velocity of 3.5 Å/ns up to a distance of 35 Å between the end of the donor strand and the anomeric oxygen of the ligand) resulted in the MM pose surviving the applied force – until the point of domain separation – in eight out of ten runs. A ninth simulation actually failed to unbind the molecule before the maximum allowed extension was reached. Given that the applied pulling forces (120 pN on average) were substantially greater to those common to atomic force microscopy experiments or exerted by the urine flow,[179] this indicated excellent stability of the MD-refined docking pose. Interestingly, during the pulling runs the LD did not show deviation toward a tighter conformation (that could further support the binding), indicating that either longer sampling times were required or that the domain separation and LD reorientation did not occur simultaneously but rather in sequence. Also, a closer analysis of the subunit movement, revealed that the extended conformations of the FimH protein (obtained toward the end of all pulling runs) did not reflect the 1KLF domain arrangement. Specifically, while a similar extent of separation could be obtained (increase of the defined distance by around 20 Å), the dihedral angle between the domains did not twist in a manner implied by the experimental structures [Figure 4.19]. While possibly caused by lack of additional domains transmitting the pulling force along the pilus rod, this could once again be interpreted through the conformational uniqueness of the FimC-stabilized constructs.

Finally, based on eight of ten simulations showing similar system behavior, a "global" PMF along the pulling path could be constructed [Figure 4.20]. As a result a steep energy increase was observed, reaching a maximum around 53 kcal/mol. Given the estimated native MM binding free energy of close to 13 kcal/mol (as obtained from MM/GBSA calculations), this value along with the shape of the curve, clearly indicated a substantial cost (around 40 kcal/mol) of the domain separation preceding ligand unbinding. In order to explore this further, a so-called "local" PMF was also plotted,[116] corresponding to free energy changes within 1 Å bins defined along the pulling path [Figure 4.20]. As a result, an upwards slope littered with multiple local minima and maxima could be observed, followed by a steep energy decrease related to loss of binding. The presence of these small barriers separating meta-stable conformations, illustrated why the relatively weakly bound ligand – and not the native inter-domain interaction – was able to withstand

the applied pulling force, and was found in good agreement with a previously proposed sequential "unzipping" model for FimH.[180] Still it should be noted, that due to an arbitrary definition of the bin size, the corresponding energy values hold little significance, making the local PMF only a qualitative illustration of the underlying free energy landscape.

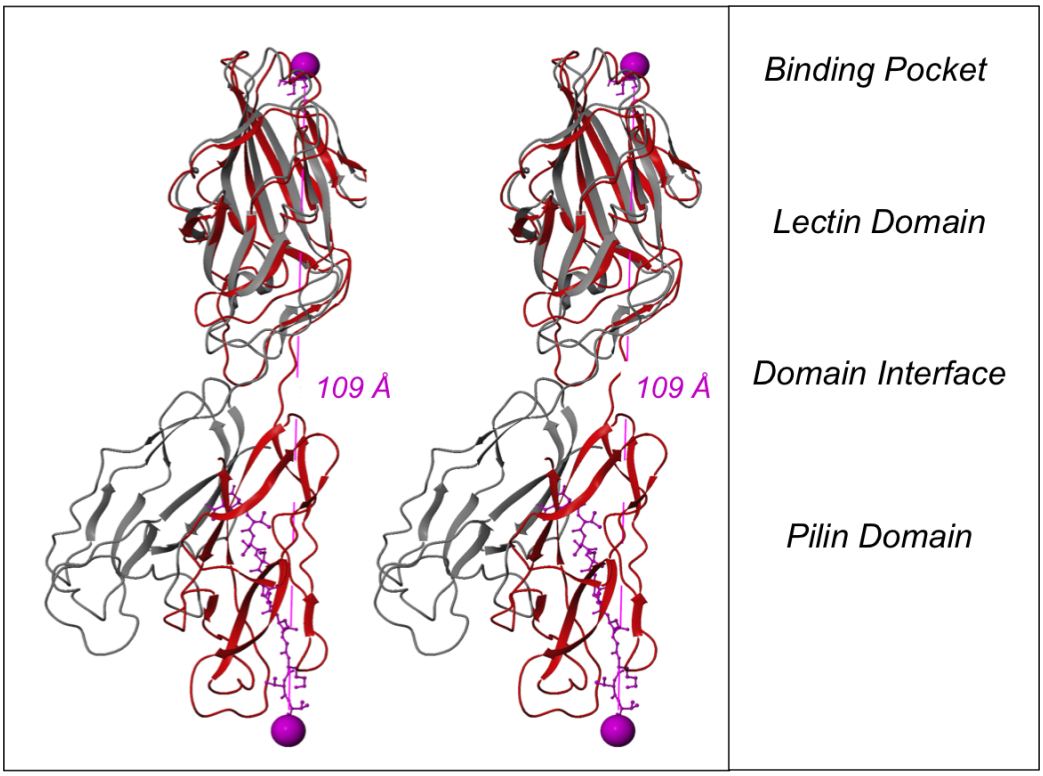


Figure 4.19: Stereo representation of the extended SMD FimH conformation (red) and the FimC-stabilized 1KLF crystal structure (gray). Purple spheres represent the atoms that were being pulled away from each other (starting distance: 81 Å). Noticeable are the differences in the LD conformations and the PD positions.

Taken together, the employed SMD protocols succeeded in simulating key stages of the bacterial cell adhesion process and confirmed the validity of the obtained MM docking pose. Still, while likely having yielded overestimated values due to high pulling velocities, they also confirmed the presence of a substantial energetical barrier related to shifting the FimH conformation between its two crystal structure states, undermining the possibility of this transition taking place under static conditions. Finally, it could be argued, that given no trace of LD conformational changes, the closure of this domain was the final stage of the catch-bonding mechanism that could only occur with the subunit separation concluded and stabilized by shear forces.

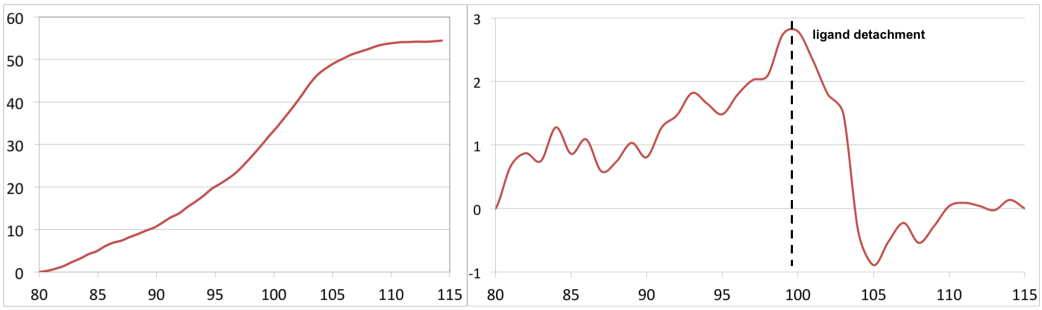


Figure 4.20: (left) Global potential of mean force, corresponding to pulling the ligand anomeric oxygen and the furtherest donor strand atom away from each other. (right) Local potential of mean force, illustrating the free energy changes within arbitrarily defined 1 Å bins. Distances given in Å and energies in kcal/mol.

4.3.4 Accelerated molecular dynamics

While the binding to the native FimH conformation – albeit requiring rearrangement within the binding pocket region – was confirmed, arguments against both discussed mechanisms of transition between known states could be raised. Particularly, the induced-fit scenario was undermined by the very large extent of the structural change – involving virtually every FimH residue – and the instability of the extended conformation. A selection mechanism on the other hand was put in doubt based on the significant energetic barriers discovered between the conformers. These findings indicated that the FimH structure presented by the majority of available crystal structures – with the LD separated from the PD, rearranged, and tightly locked around the mannosidic ligands – could not be assumed under no-flow conditions. From a computational viewpoint, definitive evidence toward/against these assumptions could be obtained from sufficiently long MD simulations of apo and holo systems. In practice however, approaching the relevant timescales (likely in the μs or ms ranges) would require months of extensive computation (100 ns MD simulations were nonetheless run for both FimH conformers, but yielded minimal change within the domains). In order to circumvent this limitation, the accelerated molecular dynamics (AMD) protocol was employed. Through this technique, we could raise the local energy minima in which the systems spend most of their time, reducing the required sampling times by a factor of 500 to 2000,[181] and enabling observation of rare events.

Due to the large and non-selective boost (strongly enhancing the motion of the natively flexible, surface-exposed moieties), initial holo simulation attempts resulted in loss of ligand binding. With the previously confirmed stability of the MM poses, this could however be safely remedied by weakly

constraining the ligand to the vicinity of the binding pockets. Following this, four 50 ns AMD simulations (theoretically corresponding to 25–100 μs of conventional MD) were run for the 1KLF- and 3JWN-based holo and apo systems discussed in previous sections. To enable analysis (beyond the level of visual inspection) focused on relevant protein backbone motion, the following metrics were employed:

- RMSD from the position of all 279 C_α 3JWN atoms, reflecting the changes in the respective domain position (difference between initial compressed and elongated conformers: 15.0 Å).
- RMSD from the position of 116 C_α 1KLF atoms outside of the domain interface, illustrating closeness to the tighter LD arrangement (difference between initial compressed and elongated conformers: 3.6 Å).
- RMSD from the position of Phe1, Ile13, Asp47, Ile52, Asp54, Gln133, Asn135, and Asp140 1KLF heavy atoms, reflecting resemblance to the rigid interior of the high-affinity pocket (difference between initial compressed and elongated conformers: 3.5 Å).

As a result, both extended FimH conformations could be seen collapsing toward a compressed 3JWN-like state, with the bound ligand proving of no observable consequence. Its presence however, could be seen to somewhat support the LD conformation, which – while still affected – would stabilize around 2 Å from the initial 1KLF arrangement. The same could be stated for the holo 1KLF binding pocket that, aside from displacement of residues Ile13 and Asp140, retained close resemblance to the starting arrangement. Interestingly, the pockets and LDs alike could also be seen populating similar conformations when both 3JWN-based simulations were run, with the RMSDs often close to 2 Å from the 1KLF template. Lastly, and perhaps most surprisingly, the collapsed 3JWN domain conformation has only proven stable in absence of the small molecular binder, suggesting an allosteric effect of the protein-ligand interface formation. Specifically, while the domains never separated in presence of MM, they could be seen moving in a scissor-like fashion, constantly decreasing or increasing the angle between them [Figure 4.21]. Given the possibility of under-sampling a larger conformational event, the two holo simulations were prolonged to 100 ns (theoretically worth up to 200 μs of conventional MD). However nothing significant could be observed, beyond the trends of collapsing the 1KLF- and retaining – albeit through an ensemble of conformations – the 3JWN-domain conformations. The above-mentioned trends are illustrated in Figures 4.22 and 4.23.

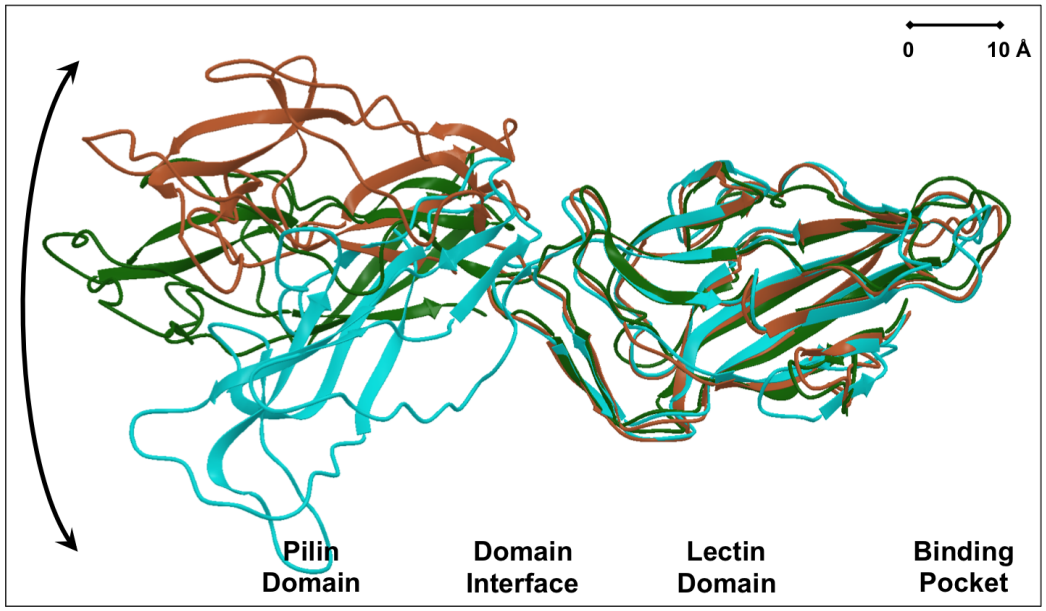


Figure 4.21: Representative structures of dominant trajectory clusters obtained from the holo 3JWN AMD simulation, superimposed on the LD C_α atoms. While a 1KLF-like domain separation was never obtained, conformations varying between a bent (blue) and nearly-linear (brown) were commonly sampled.

In summary, the AMD simulations have proven the preference of the 3JWN-like FimH conformation under static conditions, independently from the MD starting point and ligand presence. Still it bears mentioning, that the binding-site region of the native structure, was found capable of accessing a large variety of conformations. Given that this could be observed for both apo an holo simulations and that some of these conformations resembled the 1KLF arrangement, it could be hypothesized that the binding of mannosidic ligands to the F18 FimH protein was governed by a local – i.e. not involving the rest of the protein – conformational selection mechanism. Nevertheless, while strong resemblances between the sites of both states could be found – explaining the good *in vivo* potencies of ligands designed using LD-constructs –, certain residues were unable to participate in ligand binding prior to domain separation, diminishing interactions of the 2- and 3-OH mannose groups (as indicated in Figure 4.17 and reflected by the RMSD data). Specifically the previously-mentioned Ile13 residue, positioned at the tip of the "clamp loop" was never observed to near the ligand while the two FimH domains remained connected. Similarly Asp140 (expected to form a charge-assisted hydrogen bond with the 3-OH group of the mannose), would commonly maintain a distance of several ångström away from the ligand

bound to the native protein. It is assumed that the lack of these pocket elements or the cost of re-orienting them (along with their closely-neighboring residues) is primarily responsible for the differing mannosidic ligand affinities between specific protein variants and binding conditions. Finally, the previously mentioned affinity of the selective antibodies toward FimH in presence of mannose could be explained by the increased inter-domain motion likely to occasionally expose the relevant attachment region.

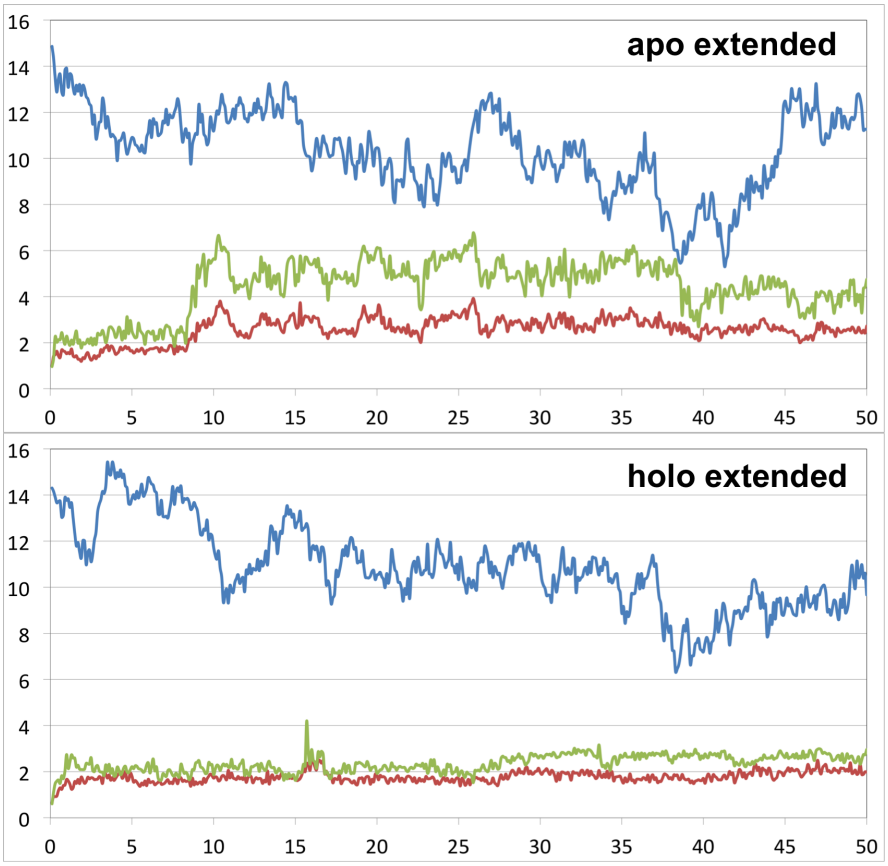


Figure 4.22: RMSDs (in Å; vertical axis) from the native compressed protein (blue), the high-affinity LD (red), and the high-affinity binding site (green). Results obtained from 50 ns apo and holo AMD simulations (horizontal axes represent the time) started from the extended FimH conformation. In both cases large domain movement could be observed, with the protein conformation slowly collapsing into the compressed arrangement. For the holo simulation, the bound ligand managed to stabilize the LD and pocket, allowing them to retain conformations close to the initial ones.

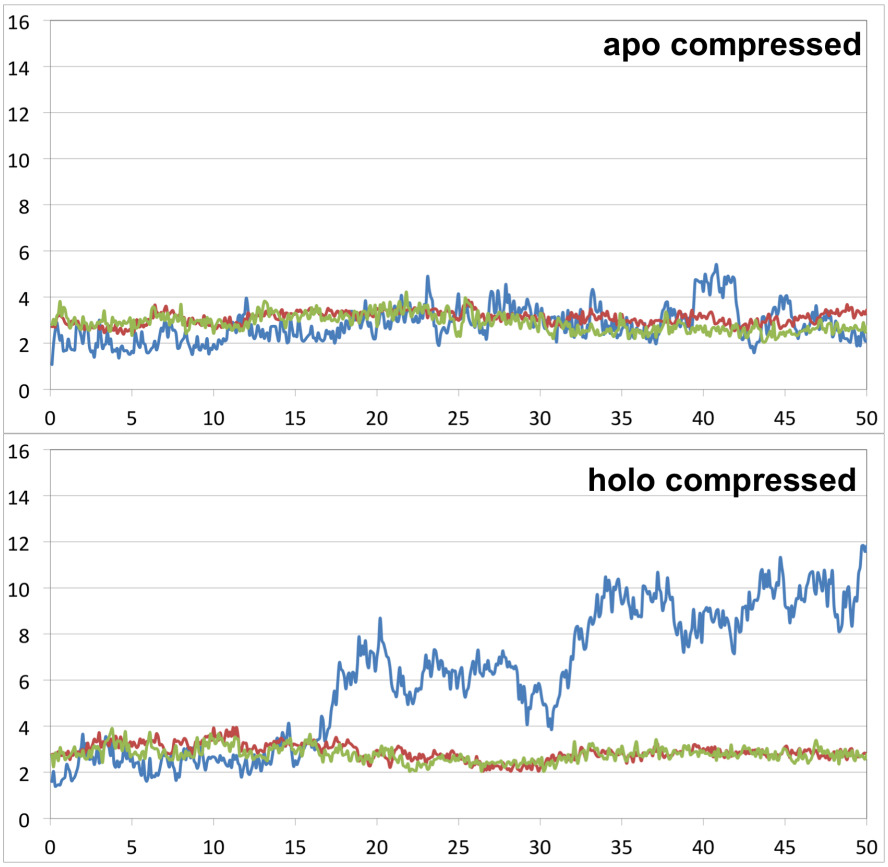


Figure 4.23: RMSDs (in Å; vertical axis) from the native compressed protein (blue), the high-affinity LD (red), and the high-affinity binding site (green). Results obtained from 50 ns apo and holo AMD simulations (horizontal axes represent the time) started from the compressed FimH conformation. While the apo protein was found relatively stable through the simulation, the inter-domain interface of the holo system was much more flexible.

4.3.5 Effects of natural mutations

Given the reported evolutionary predominance of the modeled F18 FimH variant, its compressed conformation was expected to be a valid basis for general structure-based inhibitor design. This could further be ensured through a very recent emergence of a nearly-identical K12 crystal structure,[182] (279 C_α atom RMSD of 1.2 Å between respective 3JWN and 4J3O structures). Still, the impact of the FimH sequence variations called for further probing of the relation between the naturally-occurring mutations and the affinity of mannosidic ligands. According to reported data,[88] the K12 FimH variant mediated approximately two times stronger binding of mannosidic ligands

than the F18 counterpart, while only differing by three, non-binding site residues (mutations A27V, N70S, and S78N). In light of previous observations, affinity variations such as these could relate to local movement around the compressed protein state rather than extensive structural changes. In order to confirm this and following residue modification and MD equilibration (using the 3JWN construct as basis given that the 4J3O structure was not yet available at the time), 50 ns AMD simulations were also run on apo and holo FimH systems representing the K12 variant. As a result, different microstate preferences could indeed be observed, still strongly modulated by the mannosidic ligand presence [Figure 4.24]. Nevertheless, while the K12 holo system also revealed increased inter-domain motion, its extent as well as the general resemblance to the 1KLF high-affinity pocket/LD conformation actually decreased. Even so, applying the MM/GBSA protocol on ensembles of holo AMD frames, yielded an approximately 1 kcal/mol lower energy with respect to F18, related primarily to desolvation costs. While difficult to confirm given the amount of "noise" introduced by the simulation boosts, this hinted that binding to this FimH variant either favored a different pocket conformation or required lesser desolvation and rearrangement costs to form the binding interface. With the AMD protocol indicating sensitivity to allosteric ligand and mutation effects, a systematic study introducing further FimH variants and multiple simulations per system would likely reveal more of the underlying structural characteristics.

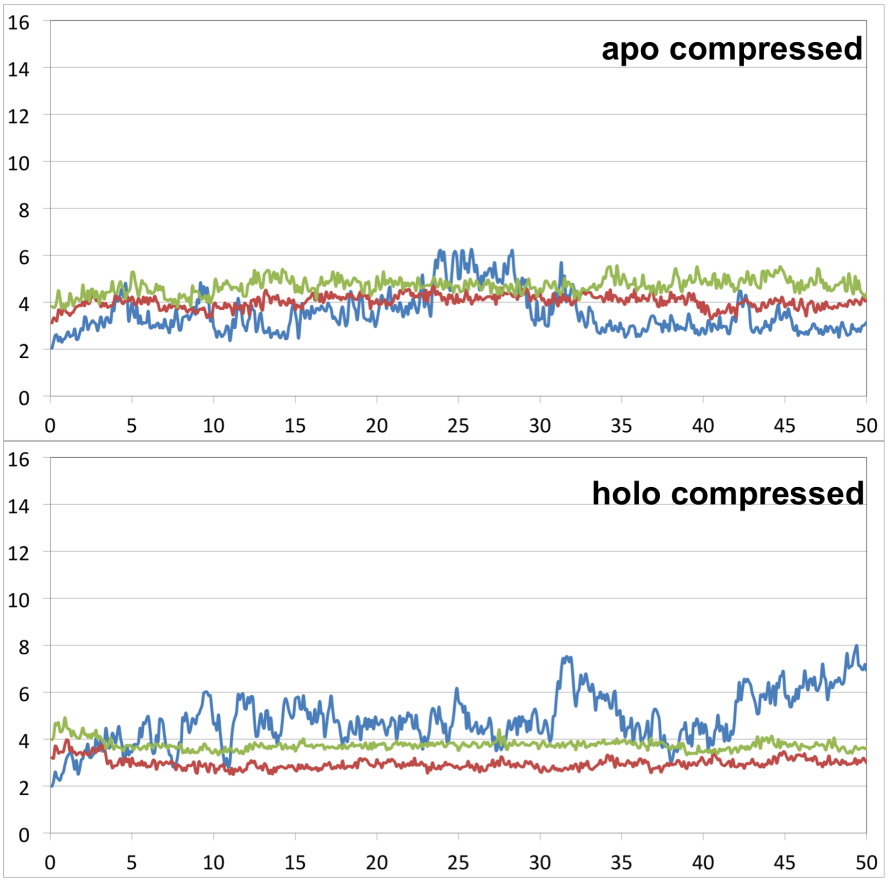


Figure 4.24: RMSDs (in Å; vertical axis) from the native compressed protein (blue), the high-affinity LD (red), and the high-affinity binding site (green). Results obtained from 50 ns apo and holo AMD simulations (horizontal axes represent the time) started from the compressed conformation of the K12 FimH variant. While the ligand presence could be seen to increase the inter-domain movement, its effect was less pronounced than for F18 FimH.

5. Conclusions and summary

While the nearly three decades of intensive research devoted to inhibiting the bacterial FimH adhesin yielded an array of compounds with excellent properties, certain structural aspects of the target itself remained elusive, impairing further progress. These involved primarily the relationships between the protein and various parts of its ligands as well as factors inducing the catch/slip-bonding behavior. Regarding the former, work described here and based on a variety of state-of-the-art *in silico* technologies, yielded a detailed description of the forces stabilizing the interface between FimH and its mannosidic ligands. While the bulk of these findings reflected the already published data, the presented optimization strategies and level of quantitative detail were – to the authors best knowledge – unprecedented. Furthermore between the modeling and previously published crystallographic studies, certain discrepancies were found, likely owing the different conditions imposed by both approaches. Of special significance, was the discovery of distinctively different characteristics of two amino acid groups partaking in the ligand binding. On one hand, the polar residues Phe1, Asp47, Asp54, Gln133, Asn135, and Asp140 within the mannose-binding pocket were confirmed responsible – through the formation of a tight hydrogen-bond network – for the bulk of ligand potency and specificity. Among these, Phe1 and Asp54 were of utmost importance, given the charge-assisted nature and excellent stability of their interactions. With the pocket perfectly pre-arranged to receive its ligands (in the case of the detached LD or FimC-stabilized complexes), this part of the interface offered an example of "traditional" binding, dominated by enthalpic components. On the other hand, a group of amino acids forming the *hydrophobic rim* at the periphery of the pocket (especially the residues Tyr48 and Tyr137 of the FimH *tyrosine gate*), defied the classical paradigms of structure-based design. Although exploring the underlying characteristics has proven challenging (as illustrated by the mutation study discussed in Section 4.1.3), evidence could be collected toward a highly-flexible and non-specifically interacting nature of this region, voiding the distinction of rigid and mutually-exclusive docking modes. Between these two "faces" of the binding site, the optimization of

interactions with the interior pocket presents – in the opinion of the author – a greater appeal. This conclusion could be reached based on the limited reach of the *tyrosine gate* and the persisting entropic phenomena preventing directional interaction with the aglycones. The mannose pocket on the other hand, was shown to tolerate modifications of the carbohydrate at the C-2 position that extended towards previously untapped regions of the protein.

If hypothetical interactions with additional residues were the sole argument for the synthetically-challenging modification of the mannose moiety, this direction would not be worth perusing. However, a further argument was presented through the advanced MD protocols employed towards studying the full, two-domain FimH. These unprecedented studies, revealed a difference between the structure and properties of the protein attached to the highly-mannosylated urothelial surface and its "inhibitor-binding" form, not subjected to shear stress. Specifically, energy scoring and SMD/AMD simulations highlighted a substantial energetical barrier separating the known crystal conformations of FimH (previously referred to as high- and low-affinity states), unlikely to be crossed without the influence of an external force. Due to such, an alternative arrangement of the ligand-binding site was proposed for the native protein conformation, in which some of the previously-assumed interactions of the 2- and 3-OH carbohydrate groups were not present or required additional residue re-orientation costs to be formed. Compounds satisfying both pocket arrangements (possibly already at our disposal thanks to the efforts discussed in Section 4.1.4), present a more versatile means of targeting the vast array of naturally-occurring FimH variants. In addition, given that favorable site arrangements were also observed during apo MD simulations with the native protein (as indicated by RMSD measurements reported in Section 4.3.4), a local conformational selection mechanism (i.e. not involving rearrangement of the rest of the protein) is proposed for FimH-mannoside binding under static conditions.

Finally, the broad range of *in silico* techniques presented in this work has proven beneficial toward the design, screening, optimization, and rationalization of FimH-mannoside binding. Nevertheless, it could also be shown that several commonly-employed technologies failed to accurately reproduce the *in vitro* affinities of the tested compound sets. The reason for that, was the importance of solvation and entropic phenomena – aspects still difficult to address with the standard methods – for the surface-exposed protein-ligand interface. Thus, with the indispensable support of experimental data, more suitable technologies could be used toward establishing the *VirtualDesignLab* platform (Section 4.2), that demonstrated significantly improved predictive capabilities. In addition, the employment of MD protocols has proven invaluable for further studying the FimH-ligand interfaces. With the bacterial

adhesin having proven in equal part fascinating and challenging, these and other powerful and inexpensive tools, stand ready to assist the work of medicinal chemists.

6. Bibliography

[1] Emil Fischer. Einfluss der configuration auf die wirkung der enzyme. *Berichte der deutschen chemischen Gesellschaft*, 27(3):2985–2993, Oct 1894.

[2] HR Bosshard. Molecular recognition by induced fit: how fit is the concept? *News Physiol Sci*, 16:171–173, Aug 2001.

[3] FA Walker. Models of the cytochromes b. 1. effect of substituents, axial ligand plane orientation, and possible axial ligand bond strain on the pyrrole proton shifts of a series of low-spin monosubstituted tetraphenylporphinatoiron (iii)-bisimidazole complexes. *Journal of the American Chemical Society*, 102(9):3254–3256, 1980.

[4] Jiyun Liu, Zhongsheng Zhang, Xiaojian Tan, Wim GJ Hol, Christophe LMJ Verlinde, and Erkang Fan. Protein heterodimerization through ligand-bridged multivalent pre-organization: Enhancing ligand binding toward both protein targets. *Journal of the American Chemical Society*, 127(7):2044–2045, 2005.

[5] DE Koshland. Application of a theory of enzyme specificity to protein synthesis. *Proc Natl Acad Sci U S A*, 44(2):98–104, Feb 1958.

[6] Matthew CJ Wilce, David M Dooley, Hans C Freeman, J Mitchell Guss, Hideyuki Matsunami, William S McIntire, Christy E Ruggiero, Katsuyuki Tanizawa, and Hiroshi Yamaguchi. Crystal structures of the copper-containing amine oxidase from arthrobacter globiformis in the holo and apo forms: implications for the biogenesis of topaquinone. *Biochemistry*, 36(51):16116–16133, 1997.

[7] B Ma, S Kumar, CJ Tsai, and R Nussinov. Folding funnels and binding mechanisms. *Protein Eng*, 12(9):713–720, Sep 1999.

[8] Kei-ichi Okazaki and Shoji Takada. Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proceedings of the National Academy of Sciences*, 105(32):11182–11187, 2008.

[9] Shu-Qun Liu, YH Xie, XL Ji, Y Tao, DY Tan, KQ Zhang, YX Fu, and PTP Kaumaya. Protein folding, binding and energy landscape: A synthesis. *Protein engineering*, pages 207–252, 2012.

[10] Markus A Lill. Efficient incorporation of protein flexibility and dynamics into molecular docking simulations. *Biochemistry*, 50(28):6157–6169, 2011.

[11] Josiah Willard Gibbs. *A method of geometrical representation of the thermodynamic properties of substances by means of surfaces*. Connecticut Academy of Arts and Sciences, 1873.

[12] Holger Gohlke and Gerhard Klebe. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angewandte Chemie International Edition*, 41(15):2644–2676, 2002.

[13] Rufus Lumry and Shyamala Rajender. Enthalpy–entropy compensation phenomena in water solutions of proteins and small molecules: a ubiquitous property of water. *Biopolymers*, 9(10):1125–1227, 1970.

[14] Stephen F Martin and John H Clements. Correlating structure and energetics in protein-ligand interactions: Paradigms and paradoxes. *Annual review of biochemistry*, 82(1), 2013.

[15] DE Leckband, F-J Schmitt, JN Israelachvili, and W Knoll. Direct force measurements of specific and nonspecific protein interactions. *Biochemistry*, 33(15):4611–4624, 1994.

[16] Hans Matter, David W Will, Marc Nazaré, Herman Schreuder, Volker Laux, and Volkmar Wehner. Structural requirements for factor xa inhibition by 3-oxybenzamides with neutral p1 substituents: combining x-ray crystallography, 3d-qsar, and tailored scoring functions. *Journal of medicinal chemistry*, 48(9):3290–3312, 2005.

[17] Tarun K Mandal and Chaitali Mukhopadhyay. Binding free energy calculations of galectin-3–ligand interactions. *Protein engineering*, 15(12):979–986, 2002.

[18] George A Jeffrey and George A Jeffrey. *An introduction to hydrogen bonding*, volume 12. Oxford University Press New York, 1997.

[19] Angelo Vedani. Yeti: An interactive molecular mechanics program for small-molecule protein complexes. *Journal of Computational Chemistry*, 9(3):269–280, 1988.

[20] Carleton R Sage, Earl E Rutenber, Thomas J Stout, and Robert M Stroud. An essential role for water in an enzyme reaction mechanism: the crystal structure of the thymidylate synthase mutant e58q. *Biochemistry*, 35(50):16270–16281, 1996.

[21] CN Pace, BA Shirley, M McNutt, and K Gajiwala. Forces contributing to the conformational stability of proteins. *FASEB J*, 10(1):75–83, Jan 1996.

[22] Jeremy M Berg, John L Tymoczko, and Lubert Stryer. Glycolysis and gluconeogenesis. 2002.

[23] Tobias Schwabe and Stefan Grimme. Theoretical thermodynamics for large molecules: walking the thin line between accuracy and computational cost. *Accounts of chemical research*, 41(4):569–579, 2008.

[24] JE Jones. On the determination of molecular fields. ii. from the equation of state of a gas. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 106(738):463–477, Oct 1924.

[25] Luis Serrano, Mark Bycroft, and Alan R Fersht. Aromatic-aromatic interactions and protein stability: investigation by double-mutant cycles. *Journal of molecular biology*, 218(2):465–475, 1991.

[26] JR Grover, EA Walters, and ET Hui. Dissociation energies of the benzene dimer and dimer cation. *Journal of Physical Chemistry*, 91(12):3233–3237, 1987.

[27] Douglas S Williamson, Jenifer Borgognoni, Alexandra Clay, Zoe Daniels, Pawel Dokurno, Martin J Drysdale, Nicolas Foloppe, Geraint L Francis, Christopher J Graham, Rob Howes, et al. Novel adenosine-derived inhibitors of 70 kda heat shock protein, discovered through structure-based design? *Journal of medicinal chemistry*, 52(6):1510–1513, 2009.

[28] Jens-Uwe Peters, Patrick Schnider, Patrizio Mattei, and Manfred Kansy. Pharmacological promiscuity: dependence on compound properties and target specificity in a set of recent roche compounds. *ChemMedChem*, 4(4):680–686, 2009.

[29] Ludwig Boltzmann. *Über die Beziehung zwischen dem zweiten Hauptsatze des mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung, respective den Sätzen über das Wärmegleichgewicht*. kk Hof-und Staatsdruckerei, 1877.

[30] Rudolf Clausius. On the motive power of heat, and on the laws which can be deduced from it for the theory of heat. poggendorff's annalen der physick. *LXXIX (Dover Reprint). ISBN 0-486-59065-8, see also Rudolf Clausius, Sadi Carnot, Baron William Thomson Kelvin, Thermodynamics: Memoirs by Carnot, Clausius, and Thomson, Nabu Press April*, 20:2010, 1850.

[31] Remo Perozzo, Gerd Folkers, and Leonardo Scapozza. Thermodynamics of protein-ligand interactions: history, presence, and future aspects. *Journal of Receptors and Signal Transduction*, 24(1-2):1–52, 2004.

[32] A Chang Chia-en, William A McLaughlin, Riccardo Baron, Wei Wang, and J Andrew McCammon. Entropic contributions and the influence of the hydrophobic environment in promiscuous protein–protein association. *Proceedings of the National Academy of Sciences*, 105(21):7456–7461, 2008.

[33] A Chang Chia-en, Wei Chen, and Michael K Gilson. Ligand configurational entropy and protein binding. *Proceedings of the National Academy of Sciences*, 104(5):1534–1539, 2007.

[34] Riccardo Baron, Piotr Setny, and J Andrew McCammon. Water in cavity- ligand recognition. *Journal of the American Chemical Society*, 132(34):12091–12097, 2010.

[35] Renxiao Wang, Yipin Lu, Xueliang Fang, and Shaomeng Wang. An extensive test of 14 scoring functions using the pdbbind refined set of 800 protein-ligand complexes. *Journal of chemical information and computer sciences*, 44(6):2114–2125, 2004.

[36] Nancy M King, Moses Prabu-Jeyabalan, Rajintha M Bandaranayake, Madhavi NL Nalam, Ellen A Nalivaika, Ays egu l O zen, Tu rkan Halilog lu, Nes e Kurt Yılmaz, and Celia A Schiffer. Extreme entropy–enthalpy compensation in a drug-resistant variant of hiv-1 protease. *ACS chemical biology*, 7(9):1536–1546, 2012.

[37] Dudley H Williams and Martin S Westwell. Aspects of weak interactions. *Chemical Society Reviews*, 27(1):57–64, 1998.

[38] G Ferenczy and GM Keseru. Enthalpic efficiency of ligand binding. *Journal of chemical information and modeling*, 50(9):1536–1541, 2010.

[39] Joel Tellinghuisen and John D Chodera. Systematic errors in isothermal titration calorimetry: Concentrations and baselines. *Analytical biochemistry*, 414(2):297–299, 2011.

[40] Torsten Schwede, Jürgen Kopp, Nicolas Guex, and Manuel C Peitsch. Swiss-model: an automated protein homology-modeling server. *Nucleic acids research*, 31(13):3381–3385, 2003.

[41] Erik Laurini, Valentina Dal Col, Maria Grazia Mamolo, Daniele Zampieri, Paola Posocco, Maurizio Fermeglia, Luciano Vio, and Sabrina Pricl. Homology model and docking-based virtual screening for ligands of the σ1 receptor. *ACS Medicinal Chemistry Letters*, 2(11):834–839, 2011.

[42] Miklos Feher and Christopher I Williams. Effect of input differences on the results of docking calculations. *Journal of chemical information and modeling*, 49(7):1704–1714, 2009.

[43] Ramy Farid, Tyler Day, Richard A Friesner, and Robert A Pearlstein. New insights about herg blockade obtained from protein modeling, potential energy mapping, and docking studies. *Bioorganic & medicinal chemistry*, 14(9):3160–3173, 2006.

[44] Gianluca Rossato, Beat Ernst, Martin Smieško, Morena Spreafico, and Angelo Vedani. Probing small-molecule binding to cytochrome p450 2d6 and 2c9: An in silico protocol for generating toxicity alerts. *ChemMedChem*, 5(12):2088–2101, 2010.

[45] Woody Sherman, Hege S Beard, and Ramy Farid. Use of an induced fit receptor structure in virtual screening. *Chemical biology & drug design*, 67(1):83–84, 2006.

[46] Guray Kuzu, Ozlem Keskin, Attila Gursoy, and Ruth Nussinov. Expanding the conformational selection paradigm in protein-ligand docking. In *Computational Drug Discovery and Design*, pages 59–74. Springer, 2012.

[47] Caterina Bissantz, Gerd Folkers, and Didier Rognan. Protein-based virtual screening of chemical databases. 1. evaluation of different docking/scoring combinations. *Journal of Medicinal Chemistry*, 43(25):4759–4767, 2000.

[48] Paul S Charifson, Joseph J Corkery, Mark A Murcko, and W Patrick Walters. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *Journal of medicinal chemistry*, 42(25):5100–5109, 1999.

[49] Edward M Engler, Joseph D Andose, and Paul VR Schleyer. Critical evaluation of molecular mechanics. *Journal of the American Chemical Society*, 95(24):8005–8025, 1973.

[50] Wendy D Cornell, Piotr Cieplak, Christopher I Bayly, Ian R Gould, Kenneth M Merz, David M Ferguson, David C Spellmeyer, Thomas Fox, James W Caldwell, and Peter A Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 1995.

[51] Xiao Zhu. Development of additive and polarizable force field parameters for polypeptides. 2012.

[52] Angelo Vedani and David W Huhta. A new force field for modeling metalloproteins. *Journal of the American Chemical Society*, 112(12):4759–4767, 1990.

[53] M Zacharias, TP Straatsma, JA McCammon, and FA Quiocho. Inversion of receptor binding preferences by mutagenesis: free energy thermodynamic integration studies on sugar binding to l-arabinose binding proteins. *Biochemistry*, 32(29):7428–7434, 1993.

[54] Bill R Miller III, T Dwight McGee Jr, Jason M Swails, Nadine Homeyer, Holger Gohlke, and Adrian E Roitberg. Mmpbsa. py: An efficient program for end-state free energy calculations. *Journal of Chemical Theory and Computation*, 8(9):3314–3321, 2012.

[55] Juan Du, Huijun Sun, Lili Xi, Jiazhong Li, Ying Yang, Huanxiang Liu, and Xiaojun Yao. Molecular modeling study of checkpoint kinase 1 inhibitors by multiple docking strategies and prime/mm–gbsa calculation. *Journal of Computational Chemistry*, 32(13):2800–2809, 2011.

[56] Paul D Lyne, Michelle L Lamb, and Jamal C Saeh. Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and mm-gbsa scoring. *Journal of medicinal chemistry*, 49(16):4805–4808, 2006.

[57] Thomas A Halgren, Robert B Murphy, Richard A Friesner, Hege S Beard, Leah L Frye, W Thomas Pollard, and Jay L Banks. Glide: a new approach for rapid, accurate docking and scoring. 2. enrichment factors in database screening. *Journal of medicinal chemistry*, 47(7):1750–1759, 2004.

[58] Matthias Rarey, Bernd Kramer, Thomas Lengauer, and Gerhard Klebe. A fast flexible docking method using an incremental construction algorithm. *Journal of molecular biology*, 261(3):470–489, 1996.

[59] Martin Alexander and BK Lustigman. Effect of chemical structure on microbial degradation of substituted benzenes. *Journal of Agricultural and Food Chemistry*, 14(4):410–413, 1966.

[60] Modesto Orozco and Rafael Franco. Computer simulated process of" lead optimization": A student-interactive program. *Journal of Chemical Education*, 67(3):232, 1990.

[61] Angelo Vedani, Max Dobler, Morena Spreafico, Ourania Peristera, and Martin Smieško. Virtualtoxlab - in silico prediction of the toxic potential of drugs and environmental chemicals: evaluation status and internet access protocol. *Altex*, 24(3):153, 2007.

[62] Angelo Vedani, Hans Briem, Max Dobler, Horst Dollinger, and Daniel R McMasters. Multiple-conformation and protonation-state representation in 4d-qsar: the neurokinin-1 receptor system. *Journal of medicinal chemistry*, 43(23):4416–4427, 2000.

[63] Angelo Vedani, Max Dobler, and Markus A Lill. Combining protein modeling and 6d-qsar. simulating the binding of structurally diverse ligands to the estrogen receptor. *Journal of medicinal chemistry*, 48(11):3700–3703, 2005.

[64] Axel T Brünger. X-ray crystallography and nmr reveal complementary views of structure and dynamics. *Nature structural biology*, 4:862, 1997.

[65] Daniel Kleppner and Robert J Kolenkow. *An introduction to mechanics*. Cambridge University Press, 2010.

[66] J Andrew McCammon. Dynamics of folded proteins. *Nature*, 267:16, 1977.

[67] Frank CJM van Veggel. Molecular dynamics and fep monte carlo studies of calix [4] arene-derived complexes of eu3+: The role of the counterions investigated. *The Journal of Physical Chemistry A*, 101(15):2755–2765, 1997.

[68] Stefan Fischer, Chandra S Verma, and Roderick E Hubbard. Rotation of structural water inside a protein: calculation of the rate and vibrational entropy of activation. *The Journal of Physical Chemistry B*, 102(10):1797–1805, 1998.

[69] Srinath Cheluvaraja and Hagai Meirovitch. Calculation of the entropy and free energy from monte carlo simulations of a peptide stretched by an external force. *The Journal of Physical Chemistry B*, 109(46):21963–21970, 2005.

[70] David E Shaw, Martin M Deneroff, Ron O Dror, Jeffrey S Kuskin, Richard H Larson, John K Salmon, Cliff Young, Brannon Batson, Kevin J Bowers, Jack C Chao, et al. Anton, a special-purpose machine for molecular dynamics simulation. In *ACM SIGARCH Computer Architecture News*, volume 35, pages 1–12. ACM, 2007.

[71] Stewart A Adcock and J Andrew McCammon. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chemical reviews*, 106(5):1589–1615, 2006.

[72] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman JC Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327–341, 1977.

[73] Andrey Milchev and Kurt Binder. Static and dynamic properties of adsorbed chains at surfaces: Monte carlo simulation of a bead-spring model. *Macromolecules*, 29(1):343–354, 1996.

[74] Shankar Kumar, John M Rosenberg, Djamal Bouzida, Robert H Swendsen, and Peter A Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992.

[75] Francesco Luigi Gervasio, Alessandro Laio, and Michele Parrinello. Flexible docking in solution using metadynamics. *Journal of the American Chemical Society*, 127(8):2600–2607, 2005.

[76] Sanghyun Park and Klaus Schulten. Calculating potentials of mean force from steered molecular dynamics simulations. *The Journal of chemical physics*, 120:5946, 2004.

[77] Sara Furlan, Giovanni La Penna, Angelo Perico, and Attilio Cesàro. Conformational dynamics of hyaluronan oligomers in solution. 3. molecular dynamics from monte carlo replica-exchange simulations and mode-coupling diffusion theory. *Macromolecules*, 37(16):6197–6209, 2004.

[78] Donald Hamelberg, Tongye Shen, and J Andrew McCammon. Phosphorylation effects on cis/trans isomerization and the backbone conformation of serine-proline motifs: accelerated molecular dynamics analysis. *Journal of the American Chemical Society*, 127(6):1969–1974, 2005.

[79] Beat Ernst and John L Magnani. From carbohydrate leads to glycomimetic drugs. *Nature Reviews Drug Discovery*, 8(8):661–677, 2009.

[80] Richard D Cummings. The repertoire of glycan determinants in the human glycome. *Molecular BioSystems*, 5(10):1087–1104, 2009.

[81] M Carmen Galan, David Benito-Alifonso, and Gregory M Watt. Carbohydrate chemistry in drug discovery. *Organic & Biomolecular Chemistry*, 9(10):3598, 2011.

[82] Zachary R Laughrey, Sarah E Kiehna, Alex J Riemen, and Marcey L Waters. Carbohydrate- π interactions: What are they worth? *Journal of the American Chemical Society*, 130(44):14625–14633, 2008.

[83] Robert J Woods, Raymond A Dwek, Christopher J Edge, and Bert Fraser-Reid. Molecular mechanical and molecular dynamic simulations of glycoproteins and oligosaccharides. 1. glycam_93 parameter development. *The Journal of Physical Chemistry*, 99(11):3832–3846, 1995.

[84] Anthony D Hill and Peter J Reilly. A gibbs free energy correlation for automated docking of carbohydrates. *Journal of computational chemistry*, 29(7):1131–1141, 2008.

[85] Meike Scharenberg, Daniela Abgottspon, Evelin Cicek, Xiaohua Jiang, Oliver Schwardt, Said Rabbani, and Beat Ernst. A flow cytometry-based assay for screening fimh antagonists. *ASSAY and Drug Development Technologies*, 9(5):455–464, 2011.

[86] Daniela Abgottspon and Beat Ernst. In vivo evaluation of fimh antagonists a novel class of antimicrobials for the treatment of urinary tract infection. *CHIMIA International Journal for Chemistry*, 66(4):166–169, 2012.

[87] Isolde Le Trong, Pavel Aprikian, Brian A Kidd, Wendy E Thomas, Evgeni V Sokurenko, and Ronald E Stenkamp. Donor strand exchange and conformational changes during e. coli fimbrial formation. *Journal of structural biology*, 172(3):380–388, 2010.

[88] Julie Bouckaert, Jenny Mackenzie, José L De Paz, Beatrice Chipwaza, Devapriya Choudhury, Anton Zavialov, Karin Mannerstedt, Jennifer Anderson, Denis Piérard, Lode Wyns, et al. The affinity of the fimh fimbrial adhesin is receptor-driven and quasi-independent of escherichia coli pathotypes. *Molecular microbiology*, 61(6):1556–1568, 2006.

[89] Chasper Puorger, Oliv Eidam, Guido Capitani, Denis Erilov, Markus G Grütter, and Rudi Glockshuber. Infinite kinetic stability against dissociation of supramolecular protein complexes through donor strand complementation. *Structure*, 16(4):631–642, 2008.

[90] Erik Hahn, Peter Wild, Uta Hermanns, Peter Sebbel, Rudi Glockshuber, Marcus Häner, Nicole Taschner, Peter Burkhard, Ueli Aebi, and Shirley A Müller. Exploring the 3d molecular architecture of escherichia coli type 1 pili. *Journal of molecular biology*, 323(5):845–857, 2002.

[91] Nathan Sharon. Bacterial lectins, cell-cell recognition and infectious disease. *FEBS letters*, 217(2):145–157, 1987.

[92] Devapriya Choudhury, Andrew Thompson, Vivian Stojanoff, Solomon Langermann, Jerome Pinkner, Scott J Hultgren, and Stefan D Knight. X-ray structure of the fimc-fimh chaperone-adhesin complex from uropathogenic escherichia coli. *Science*, 285(5430):1061–1066, 1999.

[93] Chia-Suei Hung, Julie Bouckaert, Danielle Hung, Jerome Pinkner, Charlotte Widberg, Anthony DeFusco, C Gale Auguste, Robert Strouse, Solomon Langermann, Gabriel Waksman, et al. Structural basis of tropism of escherichia coli to the bladder during urinary tract infection. *Molecular microbiology*, 44(4):903–915, 2002.

[94] Julie Bouckaert, Jenny Berglund, Mark Schembri, Erwin De Genst, Lieve Cools, Manfred Wuhrer, Chia-Suei Hung, Jerome Pinkner, Rikard Slättegård, Anton Zavialov, et al. Receptor binding studies disclose a novel class of high-affinity inhibitors of the escherichia coli fimh adhesin. *Molecular microbiology*, 55(2):441–455, 2005.

[95] Isolde Le Trong, Pavel Aprikian, Brian A Kidd, Manu Forero-Shelton, Veronika Tchesnokova, Ponni Rajagopal, Victoria Rodriguez, Gianluca Interlandi, Rachel Klevit, and Viola Vogel. Structural basis for mechanical force regulation of the adhesin fimh via finger trap-like β sheet twisting. *Cell*, 141(4):645–655, 2010.

[96] Adinda Wellens, Corinne Garofalo, Hien Nguyen, Nani Van Gerven, Rikard Slättegård, Jean-Pierre Hernalsteens, Lode Wyns, Stefan Oscarson, Henri De Greve, Scott Hultgren, et al. Intervening with urinary tract infections using anti-adhesives based on the crystal structure of the fimh–oligomannose-3 complex. *PLoS One*, 3(4):e2040, 2008.

[97] Zhenfu Han, Jerome S. Pinkner, Bradley Ford, Robert Obermann, William Nolan, Scott A. Wildman, Doug Hobbs, Tom Ellenberger, Corinne K. Cusumano, Scott J. Hultgren, and James W. Janetka. Structure-based drug design and optimization of mannoside bacterial fimh antagonists. *J Med Chem*, 53(12):4779–4792, Jun 2010.

[98] Mirja Hartmann and Thisbe K Lindhorst. The bacterial lectin fimh, a target for drug discovery–carbohydrate inhibitors of type 1 fimbriae-mediated bacterial adhesion. *European Journal of Organic Chemistry*, 2011(20-21):3583–3609, 2011.

[99] Lina M Nilsson, Wendy E Thomas, Evgeni V Sokurenko, and Viola Vogel. Elevated shear stress protects escherichia coli cells adhering to surfaces via catch bonds from detachment by soluble inhibitors. *Applied and environmental microbiology*, 72(4):3005–3010, 2006.

[100] MD Säemann, T Weichhart, WH Hörl, and GJ Zlabinger. Tamm-horsfall protein: a multilayered defence molecule against urinary tract infection. *European journal of clinical investigation*, 35(4):227–235, 2005.

[101] Steen G Stahlhut, Veronika Tchesnokova, Carsten Struve, Scott J Weissman, Sujay Chattopadhyay, Olga Yakovenko, Pavel Aprikian, Evgeni V Sokurenko, and Karen Angeliki Krogfelt. Comparative structure-function analysis of mannose-specific fimh adhesins from klebsiella pneumoniae and escherichia coli. *Journal of bacteriology*, 191(21):6592–6601, 2009.

[102] G Madhavi Sastry, Matvey Adzhigirey, Tyler Day, Ramakrishna Annabhimoju, and Woody Sherman. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of computer-aided molecular design*, pages 1–14, 2013.

[103] Hui Li, Andrew D Robertson, and Jan H Jensen. Very fast empirical prediction and rationalization of protein pka values. *Proteins: Structure, Function, and Bioinformatics*, 61(4):704–721, 2005.

[104] Matthew P Jacobson, Richard A Friesner, Zhexin Xiang, and Barry Honig. On the role of the crystal environment in determining protein side-chain conformations. *Journal of molecular biology*, 320(3):597–608, 2002.

[105] Woody Sherman, Tyler Day, Matthew P Jacobson, Richard A Friesner, and Ramy Farid. Novel procedure for modeling ligand/receptor induced fit effects. *Journal of medicinal chemistry*, 49(2):534–553, 2006.

[106] John C Shelley, Anuradha Cholleti, Leah L Frye, Jeremy R Greenwood, Mathew R Timlin, and Makoto Uchimaya. Epik: a software program for pk a prediction and protonation state generation for drug-like molecules. *Journal of computer-aided molecular design*, 21(12):681–691, 2007.

[107] Richard A Friesner, Jay L Banks, Robert B Murphy, Thomas A Halgren, Jasna J Klicic, Daniel T Mainz, Matthew P Repasky, Eric H Knoll, Mee Shelley, Jason K Perry, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *Journal of medicinal chemistry*, 47(7):1739–1749, 2004.

[108] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, 2006.

[109] DA Case, TA Darden, TE Cheatham III, CL Simmerling, J Wang, RE Duke, R Luo, RC Walker, W Zhang, KM Merz, et al. Amber 12. *University of California, San Francisco*, 2012.

[110] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79:926, 1983.

[111] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An n log (n) method for ewald sums in large systems. *The Journal of chemical physics*, 98:10089, 1993.

[112] Shūichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular physics*, 52(2):255–268, 1984.

[113] Richard J Loncharich, Bernard R Brooks, and Richard W Pastor. Langevin dynamics of peptides: The frictional dependence of isomerization rates of n-acetylalanyl-n -methylamide. *Biopolymers*, 32(5):523–535, 1992.

[114] Sanghyun Park, Fatemeh Khalili-Araghi, Emad Tajkhorshid, and Klaus Schulten. Free energy calculation from steered molecular dynamics simulations using jarzynskis equality. *The Journal of chemical physics*, 119:3559, 2003.

[115] Christopher Jarzynski. A nonequilibrium equality for free energy differences. *arXiv preprint cond-mat/9610209*, 1996.

[116] Morten Ø Jensen, Sanghyun Park, Emad Tajkhorshid, and Klaus Schulten. Energetics of glycerol conduction through aquaglyceroporin glpf. *Proceedings of the National Academy of Sciences*, 99(10):6731–6736, 2002.

[117] Donald Hamelberg, John Mongan, and J Andrew McCammon. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *The Journal of chemical physics*, 120:11919, 2004.

[118] Donald Hamelberg, César Augusto F de Oliveira, and J Andrew McCammon. Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *The Journal of chemical physics*, 127:155102, 2007.

[119] Denis Bucher, Barry J Grant, and J Andrew McCammon. Induced fit or conformational selection? the role of the semi-closed state in the maltose binding protein. *Biochemistry*, 50(48):10530–10539, 2011.

[120] Miriam Sgobba, Fabiana Caporuscio, Andrew Anighoro, Corinne Portioli, and Giulio Rastelli. Application of a post-docking procedure based on mm-pbsa and mm-gbsa on single and multiple protein conformations. *European journal of medicinal chemistry*, 2012.

[121] Hemant Kumar Srivastava and G Narahari Sastry. Molecular dynamics investigation on a series of hiv protease inhibitors: assessing the performance of mm-pbsa and mm-gbsa approaches. *Journal of chemical information and modeling*, 52(11):3088–3098, 2012.

[122] Gregory D Hawkins, Christopher J Cramer, and Donald G Truhlar. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *The Journal of Physical Chemistry*, 100(51):19824–19839, 1996.

[123] Jörg Weiser, Peter S Shenkin, and W Clark Still. Approximate atomic surfaces from linear combinations of pairwise overlaps (lcpo). *Journal of Computational Chemistry*, 20(2):217–230, 1999.

[124] Nadine Homeyer and Holger Gohlke. Free energy calculations by the molecular mechanics poisson- boltzmann surface area method. *Molecular Informatics*, 31(2):114–122, 2012.

[125] Samuel Genheden, Tyler Luchko, Sergey Gusarov, Andriy Kovalenko, and Ulf Ryde. An mm/3d-rism approach for ligand binding affinities. *The Journal of Physical Chemistry B*, 114(25):8505–8516, 2010.

[126] Junmei Wang, Paul Morin, Wei Wang, and Peter A Kollman. Use of mm-pbsa in reproducing the binding free energies to hiv-1 rt of tibo derivatives and predicting the binding mode to hiv-1 rt of efavirenz by docking and mm-pbsa. *Journal of the American Chemical Society*, 123(22):5221–5230, 2001.

[127] Matthew P Jacobson, David L Pincus, Chaya S Rapp, Tyler JF Day, Barry Honig, David E Shaw, and Richard A Friesner. A hierarchical approach to all-atom protein loop prediction. *Proteins: Structure, Function, and Bioinformatics*, 55(2):351–367, 2004.

[128] Jianing Li, Robert Abel, Kai Zhu, Yixiang Cao, Suwen Zhao, and Richard A Friesner. The vsgb 2.0 model: a next generation energy model for high resolution protein structure modeling. *Proteins: Structure, Function, and Bioinformatics*, 79(10):2794–2812, 2011.

[129] Alexey Onufriev, Donald Bashford, and David A Case. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics*, 55(2):383–394, 2004.

[130] Daniel R Roe and Thomas E Cheatham III. Ptraj and cpptraj: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation*, 2013.

[131] Jianyin Shao, Stephen W Tanner, Nephi Thompson, and Thomas E Cheatham. Clustering molecular dynamics trajectories: 1. characterizing the performance of different clustering algorithms. *Journal of Chemical Theory and Computation*, 3(6):2312–2334, 2007.

[132] Steven L Dixon, Alexander M Smondyrev, and Shashidhar N Rao. Phase: a novel approach to pharmacophore modeling and 3d database searching. *Chemical biology & drug design*, 67(5):370–372, 2006.

[133] Angelo Vedani, Morena Spreafico, Ourania Peristera, Max Dobler, and Martin Smieško. Virtualtoxlab in silico prediction of the endocrine-disrupting potential of drugs and chemicals. *CHIMIA International Journal for Chemistry*, 62(5):322–328, 2008.

[134] *VirtualToxLab Online Documentation. Available online at: http://www.biograf.ch/downloads/VirtualToxLab.pdf.*

[135] Joey W Storer, David J Giesen, Christopher J Cramer, and Donald G Truhlar. Class iv charge models: A new semiempirical approach in quantum chemistry. *Journal of computer-aided molecular design*, 9(1):87–110, 1995.

[136] K Shawn Watts, Pranav Dalal, Robert B Murphy, Woody Sherman, Rich A Friesner, and John C Shelley. Confgen: a conformational search method for efficient generation of bioactive conformers. *Journal of chemical information and modeling*, 50(4):534–546, 2010.

[137] *Smieško M. Personal communication. University of Basel; Basel, Switzerland: 2012.*

[138] Markus A Lill, Angelo Vedani, and Max Dobler. Raptor: combining dual-shell representation, induced-fit simulation, and hydrophobicity scoring in receptor modeling: application toward the simulation of structurally diverse ligand sets. *Journal of medicinal chemistry*, 47(25):6174–6186, 2004.

[139] Sameh Eid, Adam Zalewski, Martin Smieško, Beat Ernst, and Angelo Vedani. A molecular-modeling toolbox aimed at bridging the gap between medicinal chemistry and computational sciences. *International journal of molecular sciences*, 14(1):684–700, 2013.

[140] Sameh Mansour Abbas Eid. *Molecular simulations of carbohydrate-protein complexes*. PhD thesis, Group of Molecular Modeling, University of Basel, Switzerland, 2013.

[141] William Rowan Hamilton and William Edwin Hamilton. *Elements of quaternions*. London: Longmans, Green, & Company, 1866.

[142] Francis Gaudreault, Matthieu Chartier, and Rafael Najmanovich. Side-chain rotamer changes upon ligand binding: common, crucial, correlate with entropy and rearrange hydrogen bonding. *Bioinformatics*, 28(18):i423–i430, 2012.

[143] Emanuele Perola and Paul S Charifson. Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding. *Journal of medicinal chemistry*, 47(10):2499–2510, 2004.

[144] PR Andrews, DJ Craik, and JL Martin. Functional group contributions to drug-receptor interactions. *Journal of medicinal chemistry*, 27(12):1648–1657, 1984.

[145] Tobias Klein, Daniela Abgottspon, Matthias Wittwer, Said Rabbani, Janno Herold, Xiaohua Jiang, Simon Kleeb, Christine Lüthi, Meike Scharenberg, Jacqueline Bezençon, et al. Fimh antagonists for the oral treatment of urinary tract infections: from design and synthesis to in vitro and in vivo evaluation. *Journal of medicinal chemistry*, 53(24):8627–8641, 2010.

[146] Oliver Schwardt, Said Rabbani, Margrit Hartmann, Daniela Abgottspon, Matthias Wittwer, Simon Kleeb, Adam Zalewski, Martin Smieško, Brian Cutting, and Beat Ernst. Design, synthesis and biological evaluation of mannosyl triazoles as fimh antagonists. *Bioorg Med Chem*, 19(21):6454–6473, Nov 2011.

[147] Adinda Wellens, Martina Lahmann, Mohamed Touaibia, Jonathan Vaucher, Stefan Oscarson, Rene Roy, Han Remaut, and Julie Bouckaert. The tyrosine gate as a potential entropic lever in the receptor-binding site of the bacterial adhesin fimh. *Biochemistry*, 51(24):4790–4799, 2012.

[148] Jean-Christophe Mozziconacci, Eric Arnoult, Philippe Bernard, Quoc Tuan Do, Christophe Marot, and Luc Morin-Allory. Optimization and validation of a docking-scoring protocol; application to virtual screening for cox-2 inhibitors. *Journal of medicinal chemistry*, 48(4):1055–1068, 2005.

[149] Nidhi Singh and Arieh Warshel. Absolute binding free energy calculations: On the accuracy of computational scoring of protein–ligand interactions. *Proteins: Structure, Function, and Bioinformatics*, 78(7):1705–1723, 2010.

[150] Zhenfu Han, Jerome S Pinkner, Bradley Ford, Erik Chorell, Jan M Crowley, Corinne K Cusumano, Scott Campbell, Jeffrey P Henderson, Scott J Hultgren, and James W Janetka. Lead optimization studies on fimh antagonists: discovery of potent and orally bioavailable ortho-substituted biphenyl mannosides. *Journal of medicinal chemistry*, 55(8):3945–3959, 2012.

[151] Lijuan Pang, Simon Kleeb, Katrin Lemme, Said Rabbani, Meike Scharenberg, Adam Zalewski, Florentina Schädler, Oliver Schwardt, and Beat Ernst. Fimh antagonists: Structure–activity and structure–property relationships for biphenyl α-d-mannopyranosides. *ChemMedChem*, 7(8):1404–1422, 2012.

[152] Carsten Grabosch, Mirja Hartmann, Jörn Schmidt-Lassen, and Thisbe K Lindhorst. Squaric acid monoamide mannosides as ligands for the bacterial lectin fimh: covalent inhibition or not? *ChemBioChem*, 12(7):1066–1074, 2011.

[153] Carsten Grabosch, Katharina Kolbe, and Thisbe K Lindhorst. Glycoarrays by a new tandem noncovalent–covalent modification of polystyrene microtiter plates and their interrogation with live cells. *ChemBioChem*, 13(13):1874–1879, 2012.

[154] Meike Scharenberg, Oliver Schwardt, Said Rabbani, and Beat Ernst. Target selectivity of fimh antagonists. *Journal of medicinal chemistry*, 55(22):9810–9816, 2012.

[155] Jenny Berglund, Julie Bouckaert, Greve Henri De, and Stefan Knight. Anti-adhesive compounds to prevent and treat bacterial infections, March 23 2005. WO Patent App. PCT/EP2005/051,364.

[156] Minoru Ishikawa and Yuichi Hashimoto. Improvement in aqueous solubility in small molecule drug discovery programs by disruption of molecular planarity and symmetry. *Journal of medicinal chemistry*, 54(6):1539–1554, 2011.

[157] Xiaohua Jiang, Daniela Abgottspon, Simon Kleeb, Said Rabbani, Meike Scharenberg, Matthias Wittwer, Martina Haug, Oliver Schwardt, and Beat Ernst. Antiadhesion therapy for urinary tract infections a balanced pk/pd profile proved to be key for success. *Journal of Medicinal Chemistry*, 55(10):4700–4713, 2012.

[158] S Vickers, CA Duncan, I-WU Chen, A Rosegay, and DE Duggan. Metabolic disposition studies on simvastatin, a cholesterol-lowering prodrug. *Drug Metabolism and Disposition*, 18(2):138–145, 1990.

[159] Michi Ishigami, Tomoyo Honda, Wataru Takasaki, Toshihiko Ikeda, Toru Komai, Kiyomi Ito, and Yuichi Sugiyama. A comparison of the effects of 3-hydroxy-3-methylglutaryl-coenzyme a (hmg-coa) reductase inhibitors on the cyp3a4-dependent oxidation of mexazolam in vitro. *Drug metabolism and disposition*, 29(3):282–288, 2001.

[160] Stefan Siegrist. Focus on the binding pocket of uropathogenic escherichia coli - a calorimetric investigation. Master's thesis, Institute of Molecular Pharmacy, University of Basel, Switzerland, 2013.

[161] EV Sokurenko, HS Courtney, SN Abraham, P Klemm, and DL Hasty. Functional heterogeneity of type 1 fimbriae of escherichia coli. *Infection and immunity*, 60(11):4709–4719, 1992.

[162] Deborah E Leckband. Novel functions and binding mechanisms of carbohydrate-binding proteins determined by force measurements. *Current Protein and Peptide Science*, 12(8):743–759, 2011.

[163] Marcel Lindegger. Synthesis and characterization of pharmacodynamic and pharmacokinetic properties of new fimh antagonists. Master's thesis, Institute of Molecular Pharmacy, University of Basel, Switzerland, 2013.

[164] Jack D Dunitz et al. The entropic cost of bound water in crystals and biomolecules. *Science-AAAS-Weekly Paper Edition-including Guide to Scientific Information*, 264(5159):670–670, 1994.

[165] Manfred Hendlich, Andreas Bergner, Judith Günther, and Gerhard Klebe. Relibase: design and development of a database for comprehensive analysis of protein–ligand interactions. *Journal of molecular biology*, 326(2):607–620, 2003.

[166] Douglas B Kitchen, Hélène Decornez, John R Furr, and Jürgen Bajorath. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature reviews Drug discovery*, 3(11):935–949, 2004.

[167] Angelo Vedani and Max Dobler. 5d-qsar: the key for simulating induced fit? *Journal of medicinal chemistry*, 45(11):2139–2149, 2002.

[168] Angelo Vedani and Max Dobler. Multidimensional qsar: Moving from three-to five-dimensional concepts. *Quantitative Structure-Activity Relationships*, 21(4):382–390, 2002.

[169] Jürgen Schleucher. Conformational entropy and protein flexibility in drug design studied by nmr spectroscopy. 2010.

[170] Stefan Boresch, Franz Tettinger, Martin Leitgeb, and Martin Karplus. Absolute binding free energies: a quantitative approach for their calculation. *The Journal of Physical Chemistry B*, 107(35):9535–9551, 2003.

[171] Jiri Novotny, Robert E Bruccoleri, Malcolm Davis, and Kim A Sharp. Empirical free energy calculations: a blind test and further improvements to the method. *Journal of molecular biology*, 268(2):401–411, 1997.

[172] Olga Yakovenko, Shivani Sharma, Manu Forero, Veronika Tchesnokova, Pavel Aprikian, Brian Kidd, Albert Mach, Viola Vogel, Evgeni Sokurenko, and Wendy E Thomas. Fimh forms catch bonds that are enhanced by mechanical force due to allosteric regulation. *Journal of Biological Chemistry*, 283(17):11596–11605, 2008.

[173] Veronika Tchesnokova, Pavel Aprikian, Olga Yakovenko, Christopher LaRock, Brian Kidd, Viola Vogel, Wendy Thomas, and Evgeni Sokurenko. Integrin-like allosteric properties of the catch bond-forming fimh adhesin of escherichia coli. *Journal of Biological Chemistry*, 283(12):7823–7833, 2008.

[174] Pavel Aprikian, Veronika Tchesnokova, Brian Kidd, Olga Yakovenko, Vladimir Yarov-Yarovoy, Elena Trinchina, Viola Vogel, Wendy Thomas, and Evgeni Sokurenko. Interdomain interaction in the fimh adhesin of escherichia coli regulates the affinity to mannose. *Journal of Biological Chemistry*, 282(32):23437–23446, 2007.

[175] Tomasz Wlodarski and Bojan Zagrovic. Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. *Proceedings of the National Academy of Sciences*, 106(46):19346–19351, 2009.

[176] Thomas R Weikl and Carola von Deuster. Selected-fit versus induced-fit protein binding: Kinetic differences and mutational analysis. *Proteins: Structure, Function, and Bioinformatics*, 75(1):104–110, 2009.

[177] Lauren Wickstrom, Asim Okur, and Carlos Simmerling. Evaluating the performance of the ff99sb force field based on nmr scalar coupling data. *Biophysical journal*, 97(3):853–856, 2009.

[178] Lina M Nilsson, Wendy E Thomas, Evgeni V Sokurenko, and Viola Vogel. Beyond induced-fit receptor-ligand interactions: structural changes that can significantly extend bond lifetimes. *Structure*, 16(7):1047–1058, 2008.

[179] Eric Miller, Tzintzuni Garcia, Scott Hultgren, and Andres F Oberhauser. The mechanical properties of e. coli type 1 pili measured by atomic force microscopy techniques. *Biophysical journal*, 91(10):3848–3856, 2006.

[180] Pavel Aprikian, Gianluca Interlandi, Brian A Kidd, Isolde Le Trong, Veronika Tchesnokova, Olga Yakovenko, Matt J Whitfield, Esther Bullitt, Ronald E Stenkamp, Wendy E Thomas, et al. The bacterial fimbrial tip acts as a mechanical force sensor. *PLoS biology*, 9(5):e1000617, 2011.

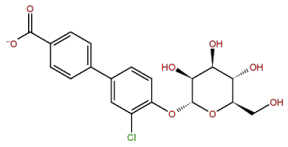
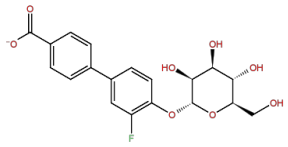
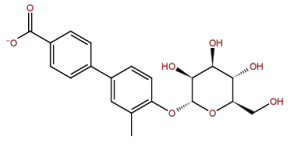
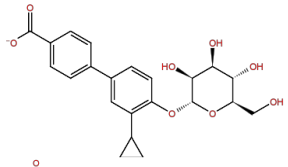
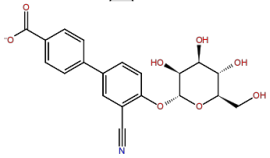
[181] Levi CT Pierce, Romelia Salomon-Ferrer, Cesar Augusto F. de Oliveira, J Andrew McCammon, and Ross C Walker. Routine access to millisecond time scale events with accelerated molecular dynamics. *Journal of Chemical Theory and Computation*, 8(9):2997–3002, 2012.

[182] Sebastian Geibel, Erik Procko, Scott J Hultgren, David Baker, and Gabriel Waksman. Structural and energetic basis of folded-protein transport by the fimd usher. *Nature*, 496(7444):243–246, 2013.

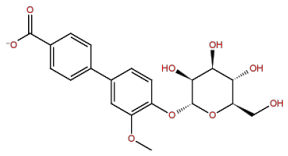
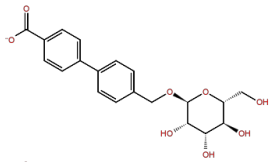
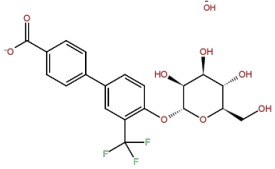
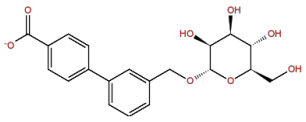
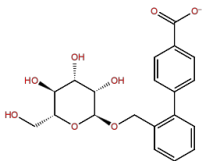
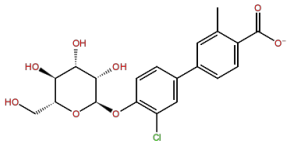
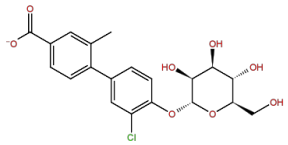
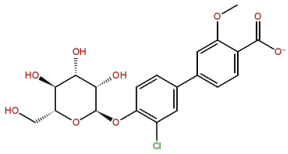
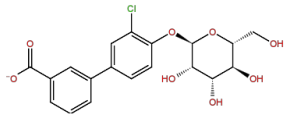
7. Appendix

7.1 Additional results (Section 4.1.3)

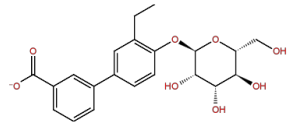
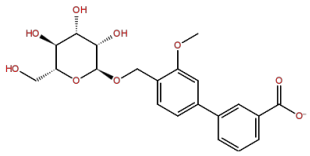
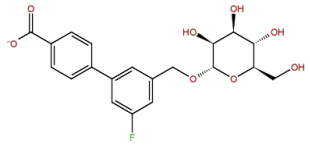
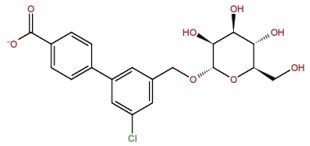
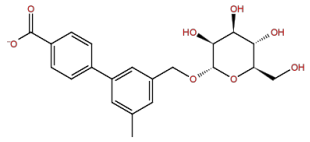
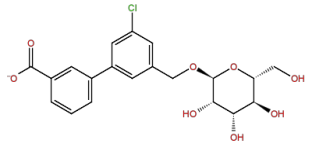
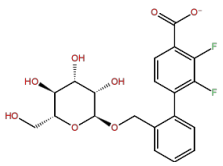
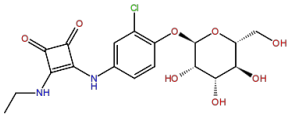
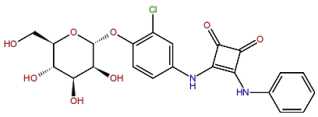
Table 7.1: Structures, docking scores (*Induced Fit Docking*), free energy estimates (MM-GBSA; software *Prime*), and ligand efficiencies of screened compounds with various aglycone modifications. All values in kcal/mol and rounded to the first decimal digit.

	Structure	Docking score	Binding energy	Ligand efficiency
1		−7.3	−92.2	−3.3
2		−6.6	−91.4	−3.3
3		−6.5	−94.3	−3.4
4		−6.8	−101.1	−3.4
5		−6.7	−93.6	−3.2

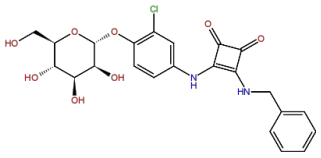
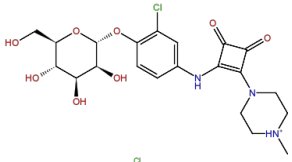
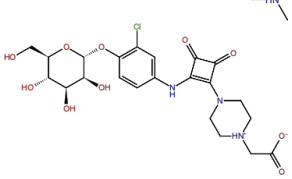
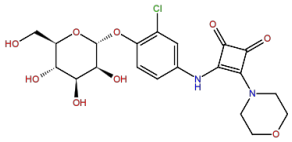
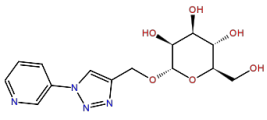
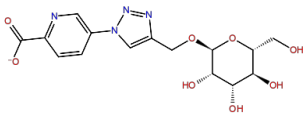
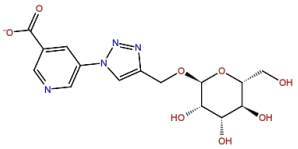
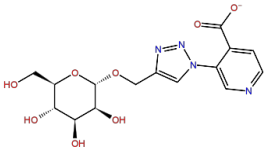
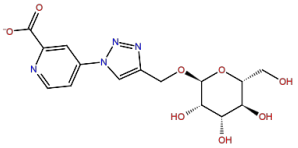
Continued from previous page ...

	Structure	Docking score	Binding energy	Ligand efficiency
6		−7.1	−95.6	−3.3
7		−6.9	−92.9	−3.3
8		−7.0	−97.3	−3.1
9		−7.3	−97.0	−3.5
10		−7.0	−97.3	−3.5
11		−7.1	−100.2	−3.5
12		−7.0	−99.7	−3.4
13		−6.9	−99.2	−3.3
14		−7.1	−98.7	−3.5

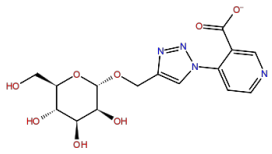
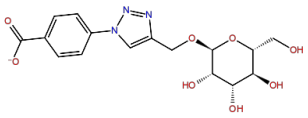
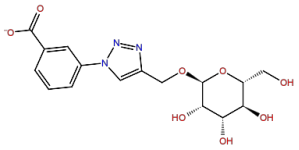
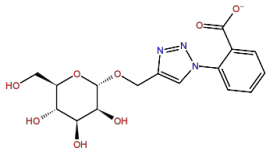
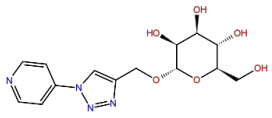
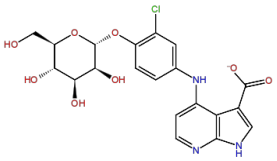
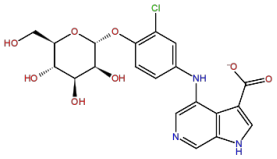
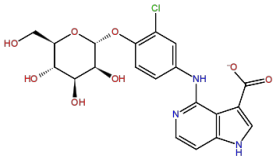
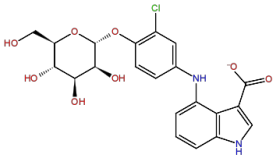
Continued from previous page ...

	Structure	Docking score	Binding energy	Ligand efficiency
15		−6.8	−95.7	−3.3
16		−6.9	−99.8	−3.3
17		−7.0	−91.3	−3.1
18		−6.8	−96.4	−3.3
19		−6.7	−93.5	−3.2
20		−7.5	−102.6	−3.5
21		−6.5	−93.9	−3.1
22		−6.7	−98.1	−3.4
23		−6.5	−106.9	−3.2

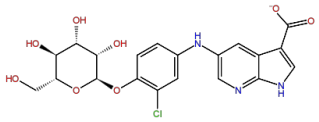
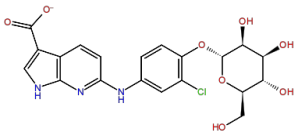
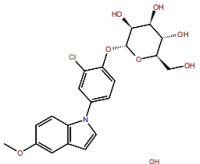
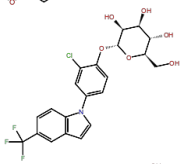
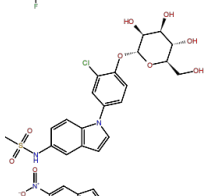
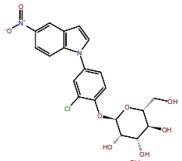
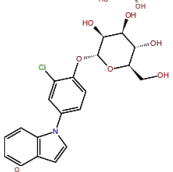
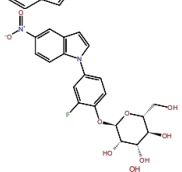
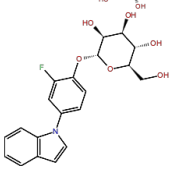
Continued from previous page ...

	Structure	Docking score	Binding energy	Ligand efficiency
24		−6.8	−104.7	−3.1
25		−6.4	−102.4	−3.1
26		−6.4	−103.7	−2.9
27		−6.3	−96.9	−3.0
28		−6.6	−82.0	−3.4
29		−5.6	−59.1	−2.2
30		−6.4	−87.7	−3.2
31		−6.4	−87.2	−3.2
32		−6.5	−88.2	−3.3

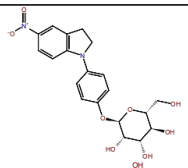
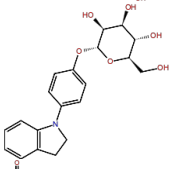
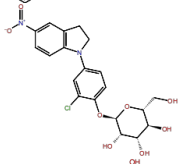
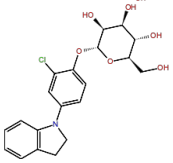
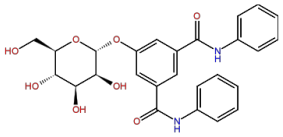
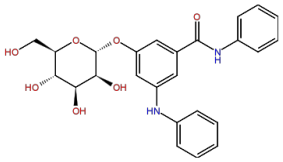
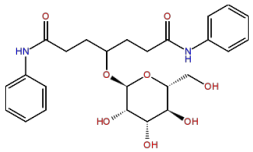
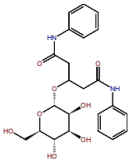
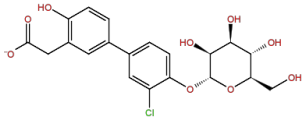
Continued from previous page ...

	Structure	Docking score	Binding energy	Ligand efficiency
33		−6.7	−88.9	−3.3
34		−6.3	−73.8	−2.7
35		−6.9	−92.6	−3.4
36		−6.4	−82.9	−3.1
37		−6.9	−89.3	−3.7
38		−6.6	−101.3	−3.2
39		−6.5	−101.1	−3.2
40		−6.3	−102.3	−3.2
41		−6.5	−95.1	−3.0

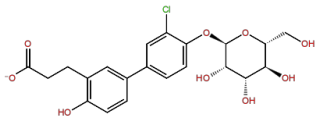
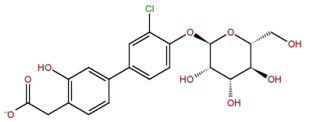
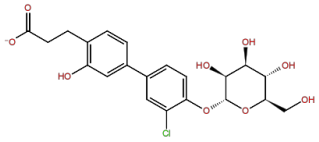
Continued from previous page ...

	Structure	Docking score	Binding energy	Ligand efficiency
42		−6.5	−94.7	−3.0
43		−7.2	−96.0	−3.0
44		−6.9	−104.4	−3.5
45		−7.1	−105.7	−3.3
46		−6.2	−102.4	−3.1
47		−7.8	−103.1	−3.3
48		−6.4	−93.2	−3.3
49		−7.5	−99.1	−3.2
50		−7.2	−101.9	−3.6

Continued from previous page ...

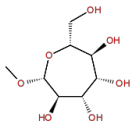
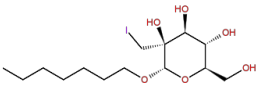
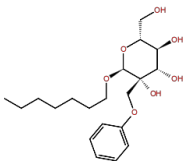
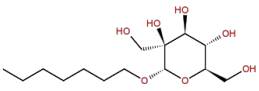
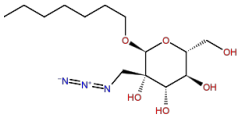
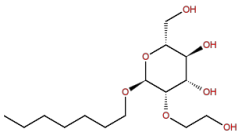
	Structure	Docking score	Binding energy	Ligand efficiency
51		−7.1	−97.5	−3.2
52		−7.2	−95.0	−3.5
53		−6.9	−103.2	−3.3
54		−7.1	−100.6	−3.6
55		−6.4	−98.9	−2.7
56		−6.6	−102.7	−3.0
57		−5.9	−112.2	−3.2
58		−6.1	−98.0	−3.0
59		−7.1	−98.3	−3.3

Continued from previous page ...

	Structure	Docking score	Binding energy	Ligand efficiency
60		−7.4	−114.8	−3.7
61		−6.7	−99.0	−3.3
62		−6.9	−100.6	−3.2

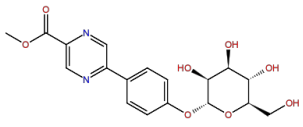
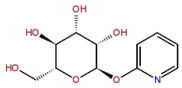
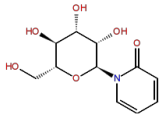
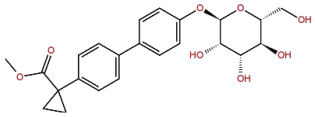
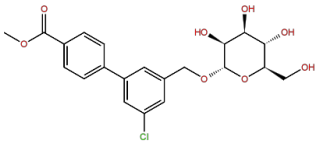
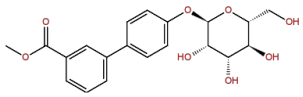
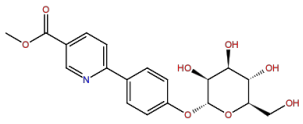
7.2 Additional results (Section 4.1.4)

Table 7.2: Structures, docking scores (*Induced Fit Docking*), free energy estimates (MM-GBSA; software *Prime*), and ligand efficiencies of screened compounds with various mannose modifications. All values in kcal/mol and rounded to the first decimal digit.

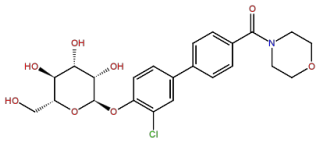
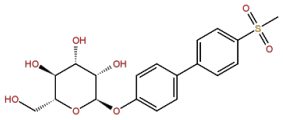
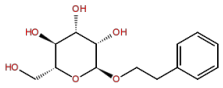
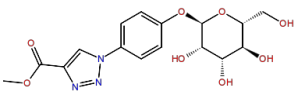
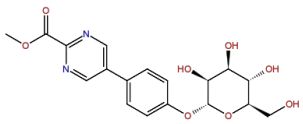
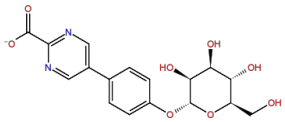
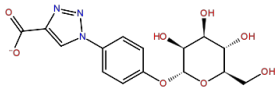
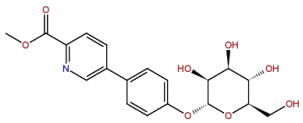
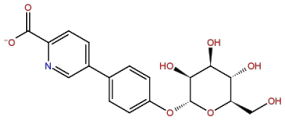
	Structure	Docking score	Binding energy	Ligand efficiency
1		−7.0	−59.0	−3.9
2		−6.5	−83.3	−4.0
3		−7.0	−92.9	−3.4
4		−6.5	−94.3	−4.5
5		−6.8	−79.8	−3.5
6		−7.1	−92.9	−4.2

7.3 Additional results (Section 4.2.2)

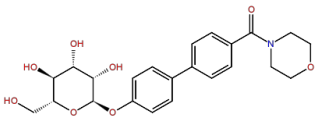
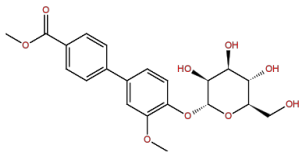
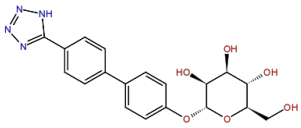
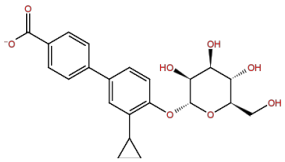
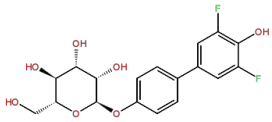
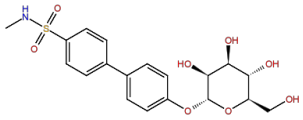
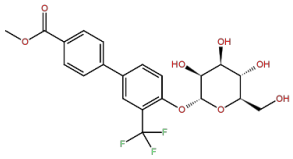
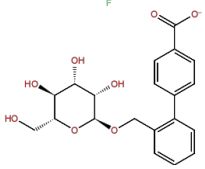
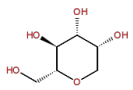
Table 7.3: Structures, docking scores (*Cheetah*), experimental free energies, and predicted free energies (*Quasar* of compounds composing the FimH QSAR model. All values in kcal/mol and rounded to the first decimal digit.

	Structure	Docking score	Exp. free energy	Predicted free energy	Residual
1		−95.1	−10.3	−10.6	0.2
2		−92.0	−8.8	−9.3	0.5
3		−82.0	−7.3	−6.7	0.6
4		−97.1	−10.1	−9.8	0.4
5		−92.9	−9.9	−9.4	0.5
6		−98.9	−10.0	−10.1	0.1
7		−92.9	−10.2	−10.0	0.2

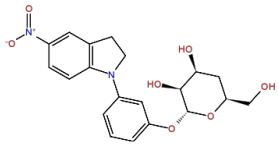
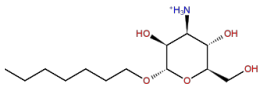
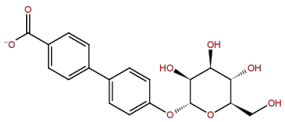
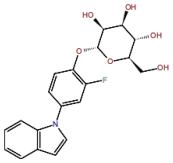
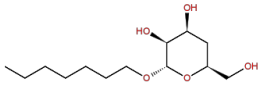
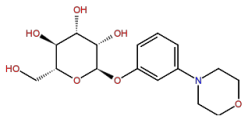
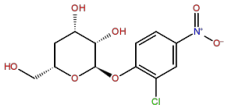
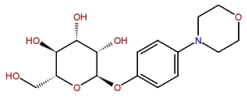
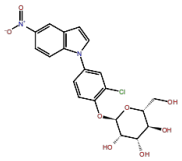
Continued from previous page ...

	Structure	Docking score	Exp. free energy	Predicted free energy	Residual
8		−98.9	−10.7	−9.9	0.8
9		−92.3	−9.5	−9.9	0.4
10		−91.2	−8.7	−9.0	0.2
11		−96.6	−10.1	−10.9	0.8
12		−91.4	−10.3	−10.7	0.4
13		−100.8	−9.2	−10.0	0.8
14		−96.7	−9.9	−10.4	0.5
15		−91.8	−9.8	−11.0	1.1
16		−103.5	−10.1	−10.4	0.4

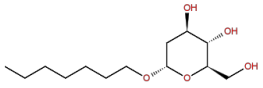
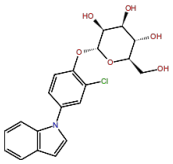
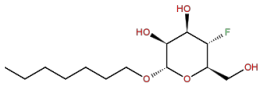
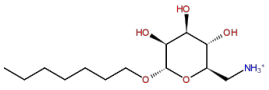
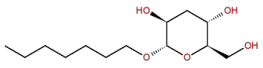
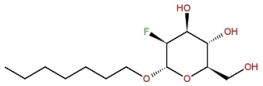
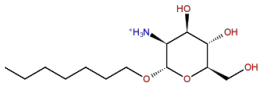
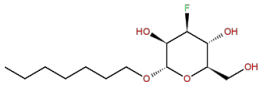
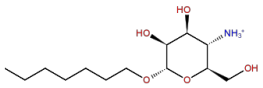
Continued from previous page ...

	Structure	Docking score	Exp. free energy	Predicted free energy	Residual
17		−96.6	−10.8	−10.6	0.2
18		−93.6	−10.3	−11.3	1.1
19		−103.6	−10.4	−10.3	0.1
20		−101.5	−10.3	−9.7	0.6
21		−95.1	−9.6	−10.3	0.7
22		−98.8	−10.8	−10.6	0.2
23		−89.8	−10.4	−11.5	1.1
24		−113.9	−9.0	−8.3	0.6
25		−88.0	−7.8	−7.5	0.3

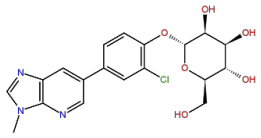
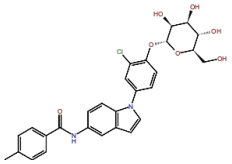
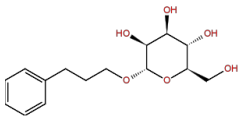
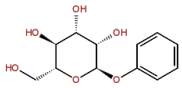
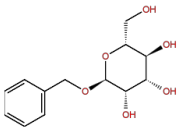
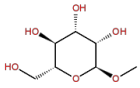
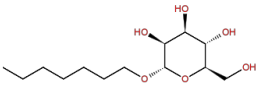
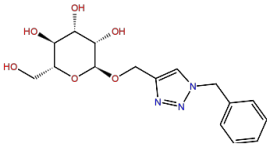
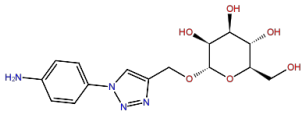
Continued from previous page ...

	Structure	Docking score	Exp. free energy	Predicted free energy	Residual
26		−81.0	−9.3	−10.5	1.2
27		−92.9	−6.4	−5.7	0.7
28		−101.5	−9.9	−10.7	0.7
29		−92.2	−9.7	−10.0	0.3
30		−75.8	−5.1	−4.9	0.2
31		−91.5	−9.7	−9.9	0.2
32		−77.1	−6.1	−5.1	1.0
33		−95.9	−9.8	−10.2	0.4
34		−90.3	−9.8	−10.5	0.7

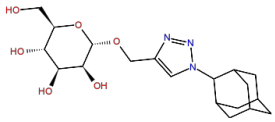
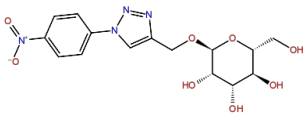
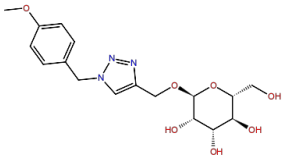
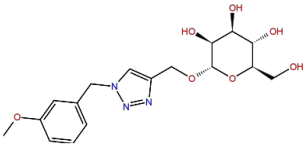
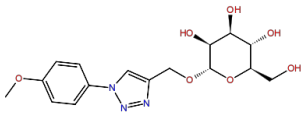
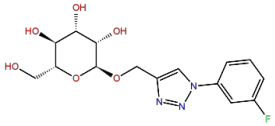
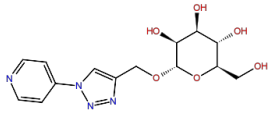
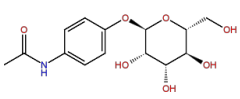
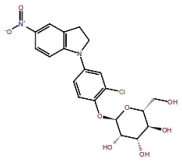
Continued from previous page ...

	Structure	Docking score	Exp. free energy	Predicted free energy	Residual
35		−75.8	−7.0	−6.7	0.4
36		−90.7	−9.3	−10.9	1.5
37		−74.6	−5.1	−5.2	0.1
38		−77.7	−6.4	−6.0	0.5
39		−73.8	−7.1	−6.0	1.1
40		−80.5	−7.4	−8.7	1.3
41		−76.3	−7.8	−8.8	1.0
42		−74.9	−7.6	−8.0	0.4
43		−104.7	−5.8	−5.4	0.4

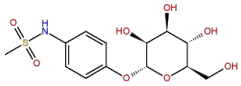
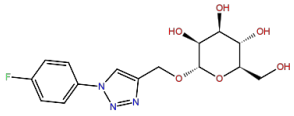
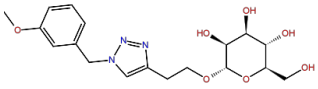
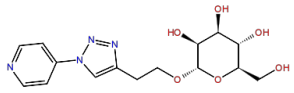
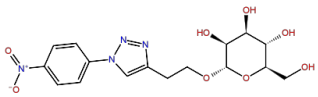
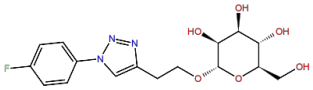
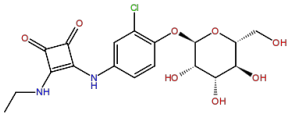
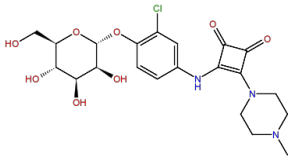
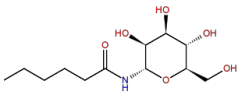
Continued from previous page ...

	Structure	Docking score	Exp. free energy	Predicted free energy	Residual
44		−94.6	−10.0	−10.5	0.4
45		−93.3	−10.6	−9.9	0.7
46		−92.3	−9.4	−8.8	0.5
47		−87.5	−8.9	−9.1	0.3
48		−88.4	−8.3	−8.6	0.3
49		−88.2	−8.0	−7.4	0.6
50		−92.2	−9.3	−9.7	0.4
51		−95.9	−9.5	−8.3	1.2
52		−94.1	−8.9	−8.5	0.4

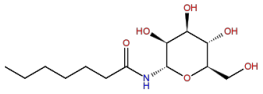
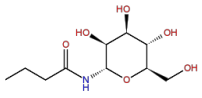
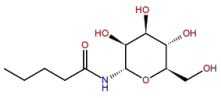
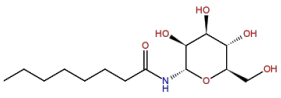
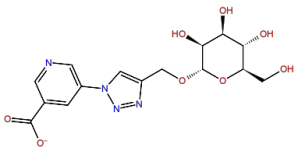
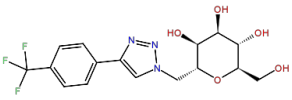
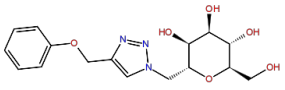
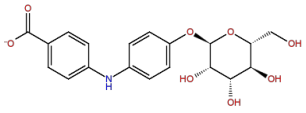
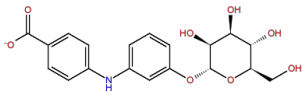
Continued from previous page ...

	Structure	Docking score	Exp. free energy	Predicted free energy	Residual
53		−92.9	−9.5	−9.2	0.2
54		−92.0	−9.5	−8.6	1.0
55		−96.9	−9.5	−8.4	1.0
56		−96.0	−9.4	−8.5	0.9
57		−94.6	−9.9	−9.1	0.8
58		−91.6	−9.0	−8.7	0.4
59		−90.8	−9.2	−9.6	0.4
60		−92.6	−9.4	−9.8	0.4
61		−93.7	−10.7	−11.6	0.9

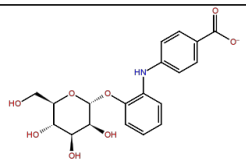
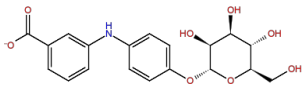
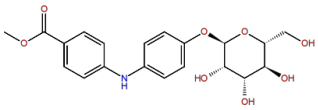
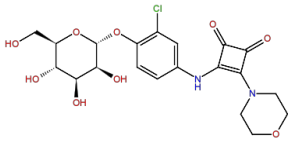
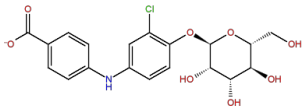
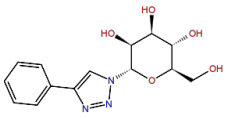
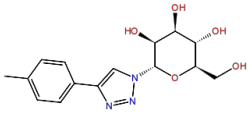
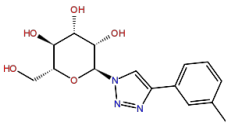
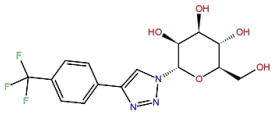
Continued from previous page ...

	Structure	Docking score	Exp. free energy	Predicted free energy	Residual
62		−102.3	−9.1	−9.0	0.1
63		−94.9	−9.1	−8.2	0.9
64		−94.0	−9.8	−8.9	0.9
65		−94.4	−9.2	−9.1	0.0
66		−93.2	−9.7	−9.3	0.4
67		−96.6	−9.3	−9.0	0.3
68		−108.8	−10.0	−10.4	0.4
69		−88.2	−10.7	−10.3	0.4
70		−87.8	−7.6	−8.8	1.2

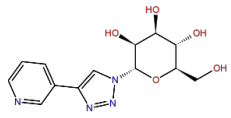
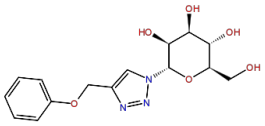
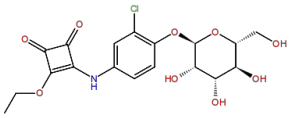
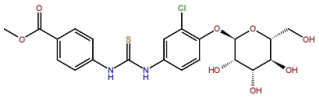
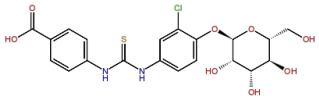
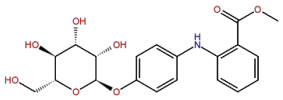
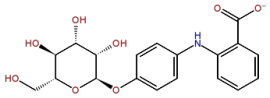
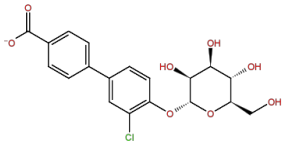
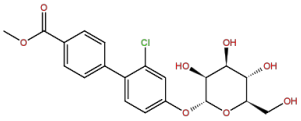
Continued from previous page ...

	Structure	Docking score	Exp. free energy	Predicted free energy	Residual
71		−90.8	−8.0	−8.9	0.9
72		−87.8	−7.9	−8.6	0.7
73		−90.5	−8.0	−8.8	0.8
74		−93.5	−8.6	−9.4	0.8
75		−100.4	−9.1	−9.1	0.1
76		−96.0	−8.9	−9.0	0.1
77		−89.2	−8.5	−8.7	0.2
78		−98.8	−10.4	−10.5	0.1
79		−94.3	−10.6	−9.8	0.8

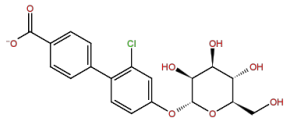
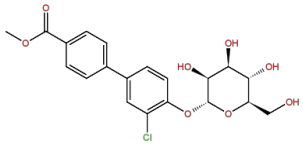
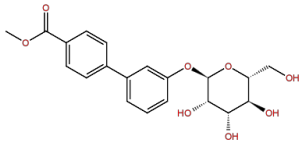
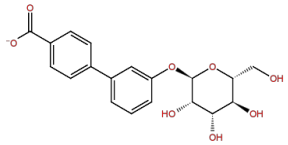
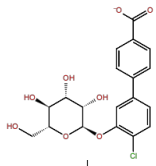
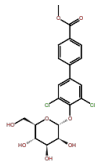
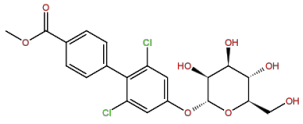
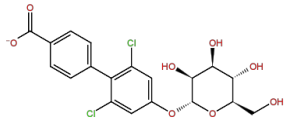
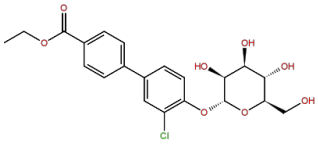
Continued from previous page ...

	Structure	Docking score	Exp. free energy	Predicted free energy	Residual
80		−110.4	−9.5	−9.6	0.1
81		−104.9	−9.8	−9.7	0.1
82		−94.2	−10.3	−10.3	0.0
83		−102.0	−10.1	−10.2	0.1
84		−102.6	−9.7	−10.4	0.7
85		−89.3	−9.1	−8.9	0.2
86		−85.3	−8.8	−8.9	0.0
87		−85.6	−9.0	−8.7	0.3
88		−89.0	−9.8	−9.2	0.6

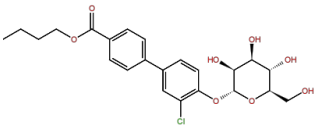
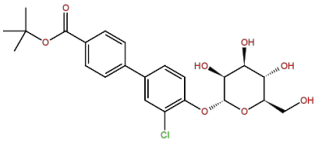
Continued from previous page ...

	Structure	Docking score	Exp. free energy	Predicted free energy	Residual
89		−92.1	−8.8	−8.9	0.2
90		−89.0	−8.7	−8.5	0.3
91		−94.9	−10.4	−10.7	0.4
92		−97.3	−9.6	−9.4	0.1
93		−95.2	−9.8	−9.4	0.4
94		−99.5	−9.4	−9.8	0.4
95		−99.5	−9.1	−9.8	0.7
96		−101.0	−10.0	−11.0	0.9
97		−89.4	−10.2	−10.3	0.0

Continued from previous page ...

	Structure	Docking score	Exp. free energy	Predicted free energy	Residual
98		−102.3	−10.0	−10.1	0.1
99		−94.9	−10.5	−11.2	0.7
100		−92.0	−9.9	−10.3	0.4
101		−93.4	−9.8	−9.9	0.2
102		−77.8	−9.8	−10.1	0.3
103		−97.0	−9.9	−10.5	0.5
104		−90.2	−10.6	−11.1	0.5
105		−101.0	−9.9	−10.7	0.8
106		−97.4	−10.5	−10.8	0.3

Continued from previous page ...

	Structure	Docking score	Exp. free energy	Predicted free energy	Residual
107		−94.5	−10.5	−9.8	0.7
108		−92.6	−10.1	−9.5	0.7

7.4 Method performance and scaling

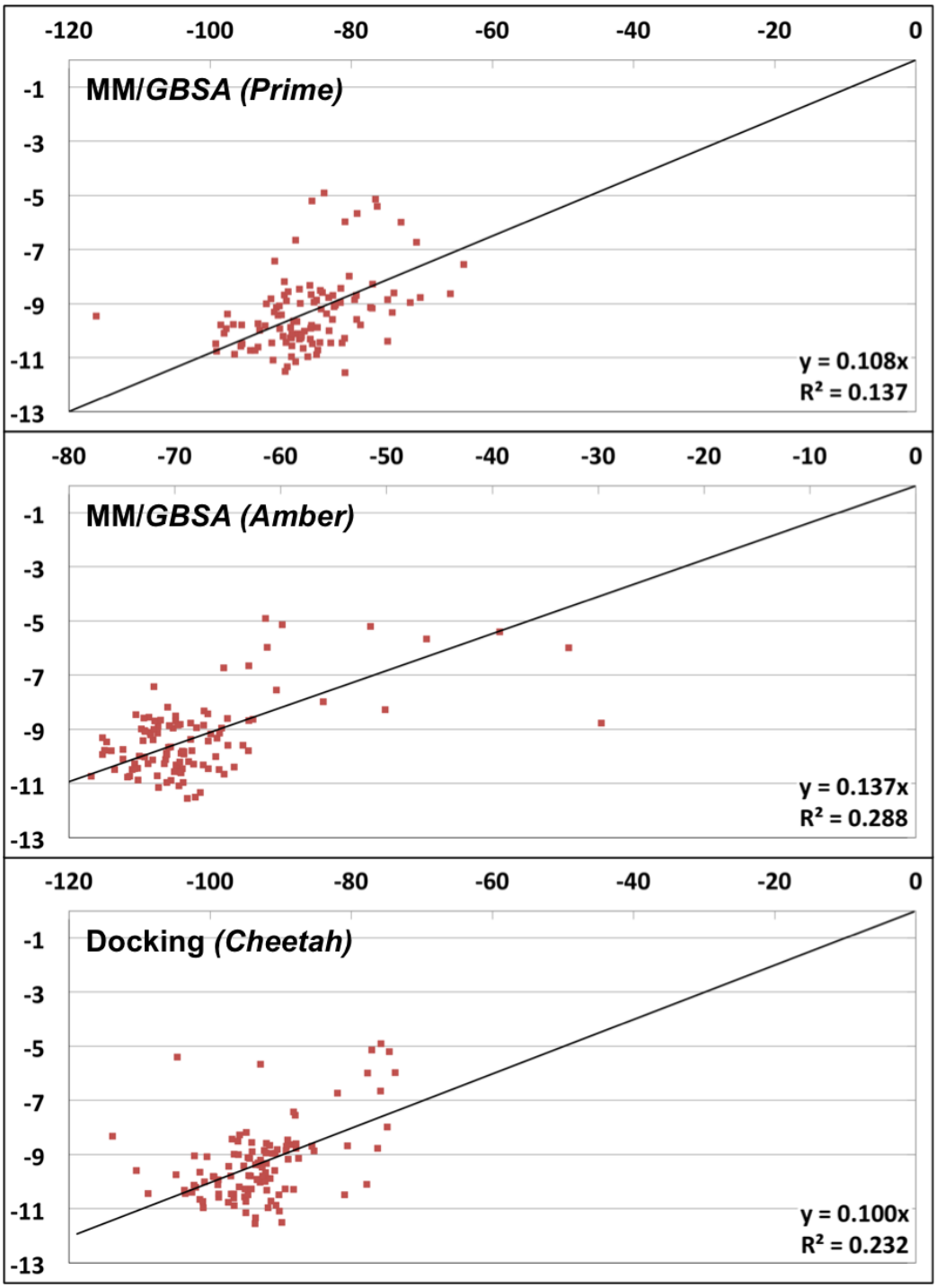


Figure 7.1: Comparison between predicted and experimental results for the 108 compounds used for developing the FimH mQSAR model. The derived scaling coefficients were 1/10, 1/7, and 1/10 for *Prime*, *Amber*, and *Cheetah* respectively. All values in kcal/mol.

7.5 *Int.J.Mol. Sci.* 2013; *14(1)*: 684-700.

Int. J. Mol. Sci. **2013**, *14*, 684-700; doi:10.3390/ijms14010684

OPEN ACCESS
International Journal of
Molecular Sciences
ISSN 1422-0067
www.mdpi.com/journal/ijms

Article

A Molecular-Modeling Toolbox Aimed at Bridging the Gap between Medicinal Chemistry and Computational Sciences

Sameh Eid [†], Adam Zalewski [†], Martin Smieško, Beat Ernst and Angelo Vedani *

Department of Pharmaceutical Sciences, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland; E-Mails: sameh.eid@unibas.ch (S.E.); adam.zalewski@unibas.ch (A.Z.); martin.smiesko@unibas.ch (M.S.); beat.ernst@unibas.ch (B.E.)

[†] These authors contributed equally to this work.

* Author to whom correspondence should be addressed; E-Mail: angelo.vedani@unibas.ch; Tel.: +41-61-267-1659; Fax: +41-61-267-1552.

Received: 12 December 2012; in revised form: 21 December 2012 / Accepted: 26 December 2012 / Published: 4 January 2013

Abstract: In the current era of high-throughput drug discovery and development, molecular modeling has become an indispensable tool for identifying, optimizing and prioritizing small-molecule drug candidates. The required background in computational chemistry and the knowledge of how to handle the complex underlying protocols, however, might keep medicinal chemists from routinely using *in silico* technologies. Our objective is to encourage those researchers to exploit existing modeling technologies more frequently through easy-to-use graphical user interfaces. In this account, we present two innovative tools (which we are prepared to share with academic institutions) facilitating computational tasks commonly utilized in drug discovery and development: (1) the *VirtualDesignLab* estimates the binding affinity of small molecules by simulating and quantifying their binding to the three-dimensional structure of a target protein; and (2) the *MD Client* launches molecular dynamics simulations aimed at exploring the time-dependent stability of ligand–protein complexes and provides residue-based interaction energies. This allows medicinal chemists to identify sites of potential improvement in their candidate molecule. As a case study, we present the application of our tools towards the design of novel antagonists for the FimH adhesin.

Keywords: computer-aided drug discovery; structure-based design; multi-dimensional QSAR; molecular dynamics; single-click molecular modeling

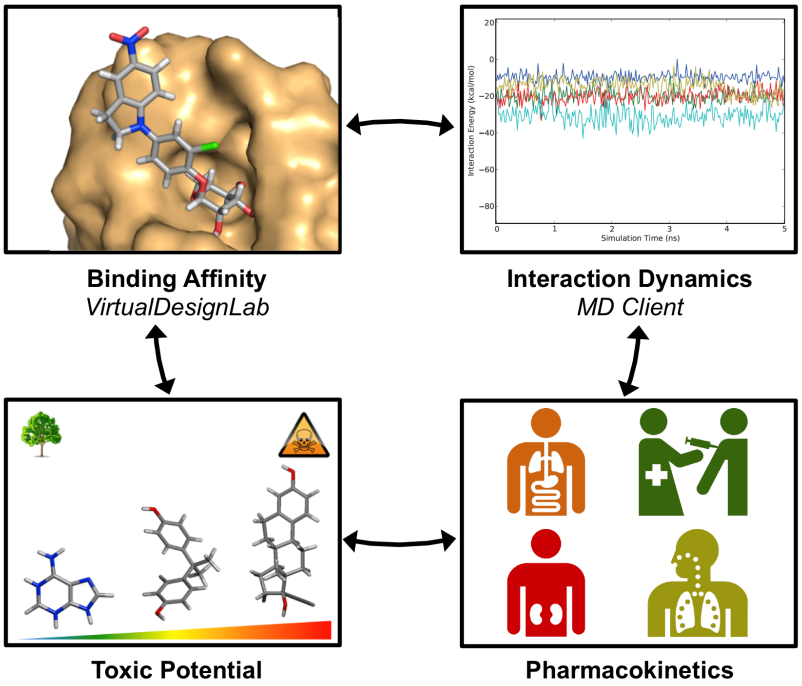
1. Introduction

Molecular modeling has become an integral part of drug discovery and development, with numerous documented examples of successful employment of computational approaches to answer key questions in the field of molecular design. For instance, structure-based design techniques, including small-molecule docking and scoring, can provide structural and energetic information on ligand–protein binding and, hence, guide the design of more potent candidate molecules [1,2]. Additionally, quantitative structure-activity relationships (QSAR) models can provide reliable estimates of binding affinities, particularly of hypothetical ligands—prior to their laborious and costly synthesis and biological testing [3,4]. Molecular dynamics (MD) simulations address more challenging questions regarding the dynamic nature of ligand–receptor interactions [5–11]. Overall, virtual screening can increase the efficiency and reduce time and cost of lead identification [12,13]. A number of commercially available software packages handle one or more of these tasks, e.g., the Schrödinger Suite [14], the Accelrys Discovery Studio [15], the SYBYL-X Suite [16] or the Molecular Operating Environment [17]. Furthermore, a wealth of modeling tools are available free-of-charge, including AutoDock for automated docking [18], Quasar^X for multi-dimensional QSAR [19], Desmond for molecular dynamics simulations [20] and DOCK Blaster for virtual screening [21].

Medicinal chemists involved in design of new ligands for some macromolecular target are nowadays knowledgeable of the binding site's topology at the molecular level. This degree of familiarity with the target provides valuable guidance for modeling techniques, such as docking proposed ligands to that target or developing a QSAR for predicting binding affinities. Such optimized modeling methodologies could, in turn, guide the medicinal chemists' decision making. However, making the best use of these and other modeling techniques requires a tedious and repetitive process of setting and calibrating parameters, as well as collecting and organizing the results. Designing an intuitive interface that encapsulates and hides the complexity of the underlying technologies from the end-user would, thus, motivate medicinal chemists to use modeling tools more frequently.

In this article, we present two novel platforms addressing commonly required tasks in modern drug design workflow: the *VirtualDesignLab* for predicting binding mode and affinity and the *MD Client* for investigating interaction dynamics of ligand–protein complexes (Figure 1). We discuss the development of the underlying models and technologies used in both tools and demonstrate their recent employment in our lab for the design and optimization of novel antagonists for FimH [22–24], a bacterial lectin playing a crucial role in the initial stages of urinary tract infections. Since the goal of the present work is to develop versatile tools that can be easily tuned for any structure-based drug design project, we will conclude with reviewing the steps required to apply/extend our tools for use with other protein targets.

Figure 1. Tools presented in this article handle two common tasks in modern computer-aided drug design workflow. The *VirtualDesignLab* predicts binding mode and estimates the associated binding affinity of prospective ligands. The *MD Client* facilitates simulation and analysis of the dynamics in ligand–protein complexes. In concert with other software predicting pharmacokinetic (e.g., QikProp [25]) and toxicological profiles (e.g., the VirtualToxLab [26]), our tools equip medicinal chemists with a multi-purpose molecular-modeling kit.



2. Methods

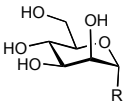
2.1. VirtualDesignLab

The *VirtualDesignLab* is an *in silico* tool developed at our institute (based on the *VirtualToxLab* framework [26] shared by the Biographics Laboratory 3R) simulating and quantifying the binding of small molecules to a macromolecular target. The technology employs automated, flexible docking combined with multi-dimensional quantitative structure-activity relationships (mQSAR). Controlled by an easy-to-use interface, the *VirtualDesignLab* allows medicinal chemists to perform quick and straightforward design, screening and structural inspection of any compound of interest [27].

In order to provide a reliable *in silico* affinity estimate for a given system, it is necessary to account for protein-ligand interactions, solvation and entropic phenomena. In our example system, FimH adhesin, we utilized a set of 108 compounds, along with their experimental affinity data, to develop and validate a corresponding mQSAR model (Table 1). When generating the model, the initial compound structures were constructed using the integrated model-building tool and then optimized with MacroModel [28]. Atomic partial charges were computed using the AMSOL package [29]. All structures were subjected to the conformational-searching algorithm ConfGen [30], resulting in sets of

low-energy conformations for each molecule in aqueous solution. Energetically feasible binding conformations (within 10 kcal/mol from the lowest-energy structure) were identified by means of automated docking to two three-dimensional structures ("in" and "out" state, cf. below) of the FimH carbohydrate-binding domain. The employed alignment (Alignator) [31] and docking (Cheetah) [32] protocols allowed for flexibility of both ligand and the protein (induced fit), as well as dynamic solvation. Several templates (based on experimental structures) were used for the pre-alignment in order to account for distinct modes of binding to FimH (referred to as "in" and "out") reported previously [23,33]. The underlying protein structures were retrieved from the Protein Data Bank (PDB codes 1UWF and 3MCY available at 1.69 Å and 2.90 Å resolution, respectively) and pre-processed (calculation of hydrogen-atom positions, hydrogen-bond network optimization, energy minimization) with the Protein Preparation Wizard in Maestro [34]. A total of 282 docking poses (allowing for multiple poses per ligand) comprising a 4D data set were then used as input (84 training and 24 test substances) for the mQSAR software Quasar [35] to generate a series of quasi-atomistic binding-site models. The underlying model families (comprising 200 members) were evaluated in consensus-scoring mode—along with a direct force-field scoring in Cheetah [32] and the comparison of a molecule's interaction energy in a box of pre-equilibrated water and in the binding site. For validation, we additionally employed an alternative receptor-modeling concept, Raptor [28], featuring a substantially different scoring function.

Table 1. Structures and binding affinities (pIC50: negative logarithm of IC_{50} [M]) for 52 compounds employed to develop the QSAR model. The remaining data cannot be disclosed at this time, due to pending patent applications.



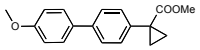
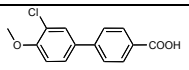
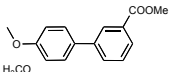
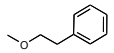
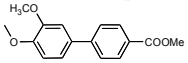
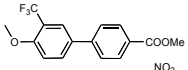
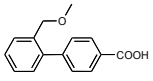
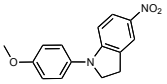
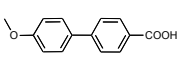
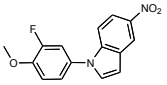
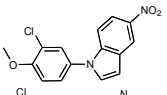
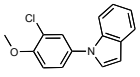
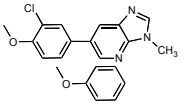
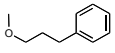
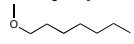
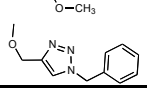
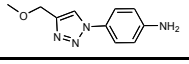
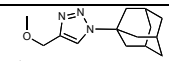
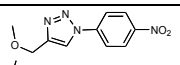
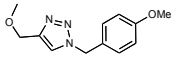
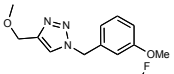
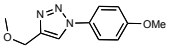
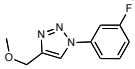
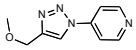
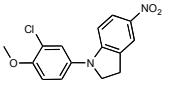
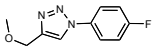
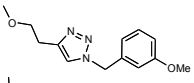
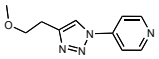
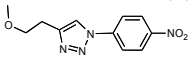
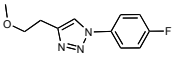
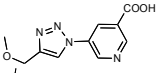
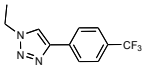
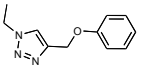
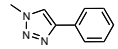
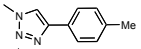
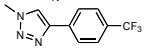
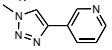
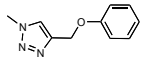
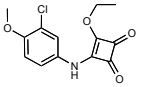
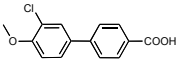
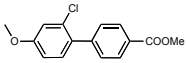
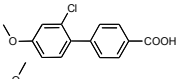
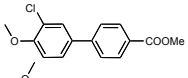
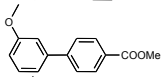
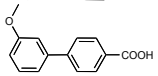
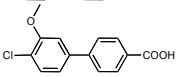
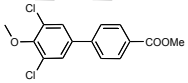
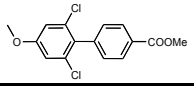
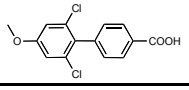
	R	Exp. affinity	Pred. affinity	Residual		R	Exp. affinity	Pred. affinity	Residual
1		7.3	7.6	0.3	**2**		7.0	7.4	0.4
3		7.5	7.5	0	**4**		6.7	6.5	−0.2
5		8.5	7.7	−0.8	**6**		8.6	7.8	−0.8
7		6.2	6.7	0.5	**8**		7.8	7.0	−0.8
9		8.0	7.4	−0.6	**10**		7.5	7.2	−0.3
11		7.8	7.3	−0.5	**12**		8.1	7.0	−1.1
13		7.8	7.5	−0.3	**14**		6.6	7.0	0.4
15		6.8	6.6	−0.2	**16**		6.4	6.2	−0.2
17		5.5	6.0	0.5	**18**		7.2	6.9	−0.3
19		6.2	7.1	0.9	**20**		6.4	6.7	0.3

Table 1. *Cont.*

	R	Exp. affinity	Pred. affinity	Residual		R	Exp. affinity	Pred. affinity	Residual
21		6.9	7.1	0.2	**22**		6.4	7.1	0.7
23		6.3	7.1	0.8	**24**		6.3	7.0	0.7
25		6.8	7.4	0.6	**26**		6.5	6.7	0.2
27		7.2	6.8	−0.4	**28**		8.6	8.0	−0.6
29		6.1	6.8	0.7	**30**		6.6	7.3	0.7
31		6.8	6.8	0	**32**		7.0	7.2	0.2
33		6.7	6.9	0.2	**34**		6.8	6.8	0
35		6.7	6.7	0	**36**		6.5	6.3	−0.2
37		6.6	6.8	0.2	**38**		6.6	6.6	0
39		6.8	7.3	0.5	**40**		6.7	6.5	−0.2
41		6.3	6.5	0.2	**42**		8.0	7.7	−0.3
43		8.2	7.5	−0.7	**44**		7.7	7.6	−0.1
45		7.6	7.5	−0.1	**46**		8.3	7.8	−0.5
47		7.7	7.4	−0.3	**48**		7.4	7.3	−0.1
49		7.5	7.3	−0.2	**50**		7.8	7.4	−0.4
51		8.3	7.9	−0.4	**52**		8.0	7.4	−0.6

Every compound submitted to the *VirtualDesignLab* server (by means of imported PDB files or the integrated model builder) is subjected to identical protocols as those employed to train and validate the underlying mQSAR model(s) (Figure 2). The affinity is calculated based on multiple components of the binding energy (Figure 3). Protein–ligand interaction and internal strain energies (Cheetah and Quasar) are obtained using a directional force field with polarization terms [36]. The desolvation costs are calculated for the global minimum obtained from the conformational search, using a continuum solvation model. Loss of entropy is approximated from the number of rotatable bonds constrained upon binding to the protein. Induced-fit energy calculation is an inherent function of the Quasar algorithm. The affinity predictions are based on (up to) eight docking poses as obtained from Alignator/Cheetah (4D) and take into account (up to) six induced-fit mechanisms (5D) and two solvation (6D) scenarios in order to account for the unique properties of certain binding sites (e.g., the surface-exposed FimH binding pocket). Protein–ligand structures may be viewed (binding pocket)

and/or downloaded (in PDB format) upon job completion. The latter files also serve as input for other software, including the *MD Client*.

Figure 2. *VirtualDesignLab* flowchart (left) and an example of a typical workflow based on the FimH receptor (right). The compound of interest is designed using the built-in 3D model builder or imported from an external file. The main step involves the conformational sampling of the ligand in the protein's binding pocket, where all feasible poses are retained and used as input for the subsequent mQSAR. Structure management options and job controls are all accessible from a central interface window. References to the individual pieces of software are given in text.

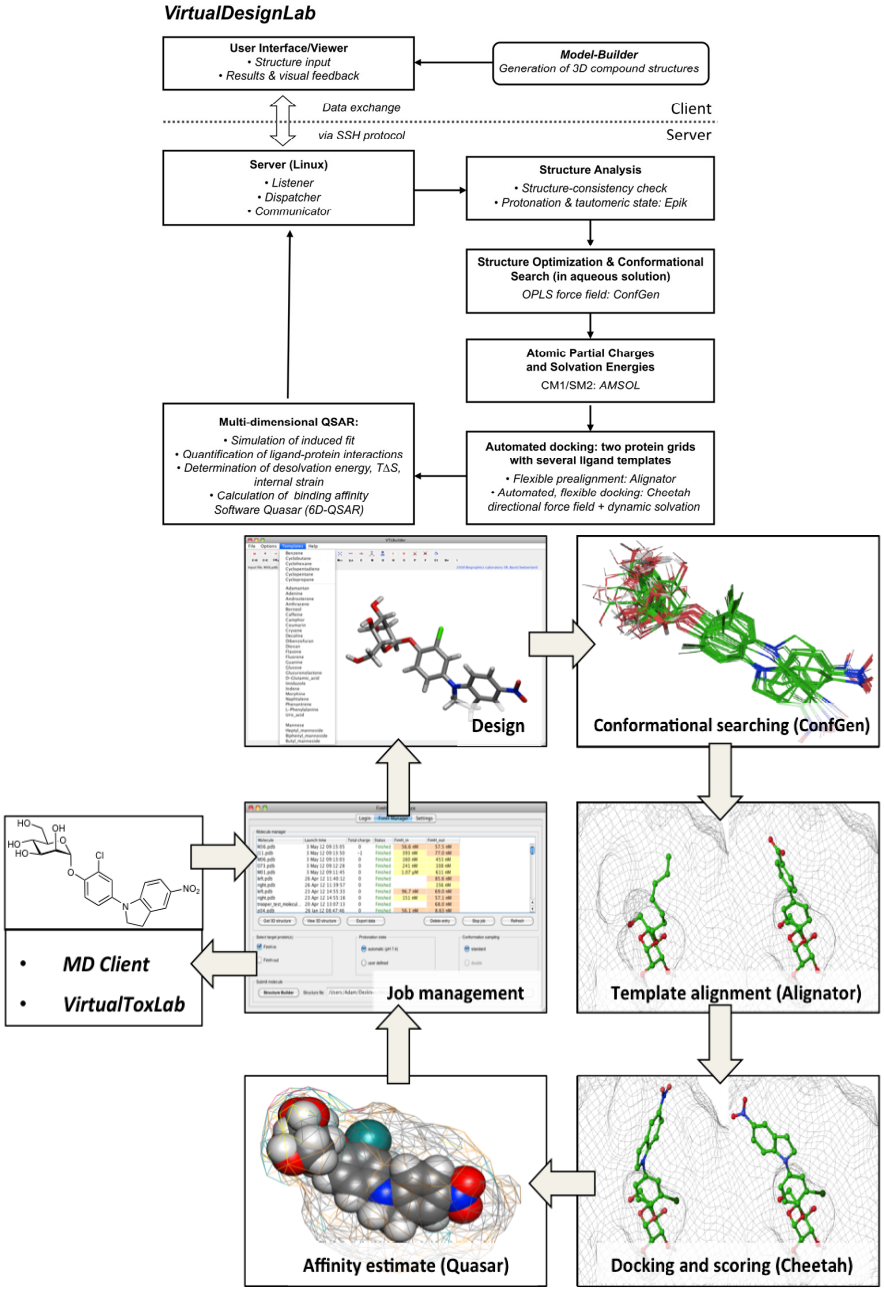


Figure 3. The equation for calculating the binding energy used in the *VirtualDesignLab/VirtualToxLab* and the directional force field employed in Cheetah and Quasar [32]. The individual terms—quantifying experimentally accessible quantities, such as bond lengths, bond angles, torsion angles, van der Waals contacts, geometries of hydrogen bonds, electrostatic and metal-ligand interactions, as well as ligand→protein polarization—are described in greater detail in the software documentation found at http://www.biograf.ch/index.php?id=software.

$$E_{binding} = E_{ligand\text{-}receptor} - E_{ligand\ desolv.} - T\Delta S - E_{ligand\ strain} - E_{induced\ fit}$$

$$E_{ligand-receptor} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{torsions} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{nb\ pairs} \frac{q_i \cdot q_j}{4\pi\varepsilon_0 D(r) r_{ij}} + \sum_{nb\ pairs} \left(\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^{6}}\right) +$$

$$\sum_{H\ bonds} \left(\frac{C}{r_{ij}^{12}} - \frac{D}{r_{ij}^{10}}\right) \cdot \cos^2(\theta_{Don-H\cdots Acc}) \cdot \cos^n(\omega_{H\cdots Acc-LP}) + \sum_{metal\ pairs} \frac{q_i^{CT} \cdot q_j^{CT}}{4\pi\varepsilon_0 D(r) r_{ij}} + \sum_{metal\ pairs} \left(\frac{E}{r_{ij}^{12}} - \frac{F}{r_{ij}^{10}}\right) + \sum_{atoms} -\frac{1}{2}\alpha_i \left[\vec{E}_i^{\circ} \cdot \vec{E}_i\right]$$

2.2. MD Client

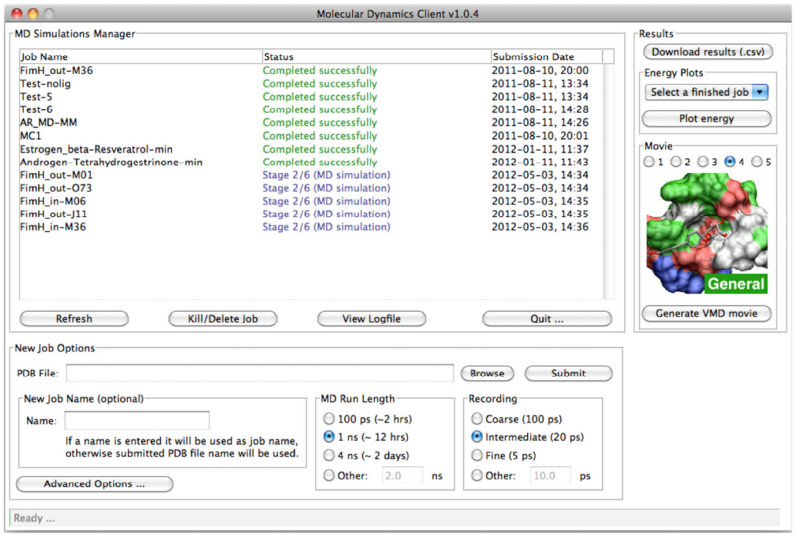
In addition to binding affinities estimated from mQSAR based on the docking simulations, medicinal chemists might wish to analyze the kinetic stability of ligand–protein interaction used by means of molecular-dynamics (MD) simulations. MD has been successful in studying structural fluctuations in proteins [37–39], lipids [40–42] and nucleic acids [43,44], as well as in the refinement of structures solved by X-ray crystallography and NMR [5]. Despite the availability of a wealth of software packages for performing MD simulations, (e.g., Desmond [20], Amber [45], CHARMM [46], GROMOS [47] and GROMACS [48]), the lengthy setup and laborious post-processing act as a barrier, preventing users from routinely utilizing these simulations. We therefore developed the *MD Client* to overcome this limitation by requiring as few settings as possible to quickly and reliably highlight basic features of the dynamics of the studied protein–ligand complex. Our *MD Client* is designed specifically for use by bench medicinal chemists interested in exploring ligand–protein interaction dynamics.

2.2.1. The *MD Client* Interface

The *MD Client* utilizes a simple and intuitive GUI front-end and a more sophisticated back-end that handles all "under-the-hood" tasks, from cleaning the input structure to post-processing MD trajectory and gathering energy results. Both front- and back-end programs were developed in python 2.6 (http://www.python.org) using standard extensions, such as the *TkInter* GUI package and the *matplotlib* library for rendering interactive 2D plots. The front-end has been compiled for Mac OS X, Linux and Windows operating systems. The communication between the front-end and the back-end on the remote server is carried out via a Secure Shell (SSH) protocol. A molecular dynamics simulation of a ligand–protein complex (as obtained, for instance, from the *VirtualDesignLab*) is launched by a single-click in the *MD Client* interface. The *MD Client* provides control over the basic parameters of submitted MD simulation: namely its length and frequency of taking snapshots for subsequent energy analysis and movie production (Figure 4). The "Advanced Options", button enables the user to control more details of the MD simulation; yet the default options are adequate in most

cases. A list box keeps track of jobs currently on the server and their current status. The user can monitor the progress of running jobs or, when needed, terminate them at any stage.

Figure 4. Appearance of the user interface of *MD Client*; top-left: list of jobs currently on remote server; bottom: basic simulation settings; and right: results download and analysis.



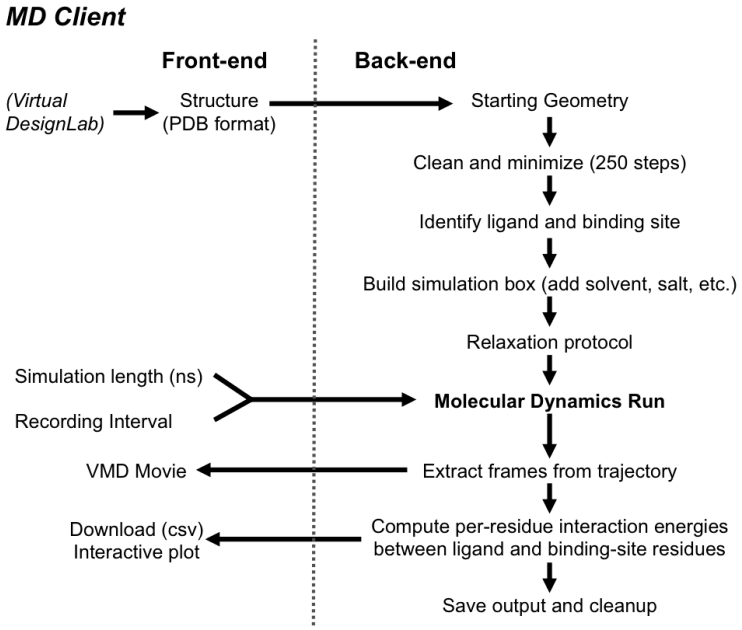
MD simulations require different types of data input. The most important is the structure file containing input geometries of the ligand–protein complex. Currently, the *MD Client* accepts files with Protein Data Bank (PDB) format [49]. The *VirtualDesignLab* output structures (*i.e.*, ligand-protein complexes) can be directly used as input for the *MD Client*. When an MD simulation is completed, the user can download extracted frames as standard PDB files for viewing. For the 3D visualization of structures, several free 3D-rendering tools, such as Visual Molecular Dynamics (VMD) [50], are available. To facilitate the importing of MD trajectories into VMD, we added a functionality that automatically generates a VMD visualization-state file linked to the downloaded frames. The user can choose between different pre-defined visualization styles and simply click on "Generate VMD Movie" in the *MD Client* interface (Figure 4) to produce a file that can be loaded directly into VMD. This spares the user the time and effort needed to load individual PDB files and set up the view options in VMD. Most importantly, the user can use *MD Client's* built-in plotting tool to analyze ligand–protein interaction dynamics (cf. Figure 7) and compare them amongst multiple systems, which we are going to demonstrate in the results section on selected antagonists binding to the FimH receptor.

2.2.2. The *MD Client* Back-end

The *MD Client* back-end resides on the remote server, where all computational jobs are to take place. It utilizes Schrödinger's Python API (http://www.schrodinger.com/pythonapi) for reading structures, launching MD simulations and computing per-residue interaction energies. It receives input structure and primary simulation settings from the front-end. Figure 5 shows how the *MD Client* back-end processes input structures into useful quantities. It starts by constructing atom connectivity and bond orders for the submitted structure and doing a short energy minimization to relieve structural

inconsistencies in bond lengths, angles, steric clashes, *etc.* The *MD Client* back-end automatically identifies the ligand-like molecule and defines binding site residues as all residues within 8 Å (default) from ligand atoms. This definition is employed for subsequent use in energy computations and movie production.

Figure 5. Workflow of the *MD Client* and communication between front-end and back-end.



The *MD Client* back-end employs the Desmond package from the D. E. Shaw Research laboratory to perform the MD simulations [51,52]. Desmond and its source code are distributed under free license to non-commercial and academic users. It uses novel parallel algorithms and numerical techniques to achieve high performance and accuracy on platforms containing a large number of processors, but may also be executed on a single-processor computer [20]. Desmond's System Builder soaks the submitted ligand–protein complex into a TIP3P water box extending 10 Å beyond any of the complex's atoms. It adds counter ions to neutralize the simulation box and 0.15 M sodium and chloride ions to approximate physiological conditions. The complex is first minimized to a convergence gradient threshold of 1.0 kcal/(mol·Å). The molecular-dynamics protocol utilizes the OPLS2005 force field and the NPT ensemble (constant number of particles, pressure and temperature) at 300 K, with periodic boundary conditions. The production run of the user-defined length is preceded by 24 ps of the Desmond default relaxation protocol. After completion of the MD simulation, the *MD Client* back-end extracts frames at the user-defined intervals and saves them as standard PDB files. The user can download these frames for later viewing and analysis. Finally, the extracted frames are analyzed using the component-interactions script in Maestro [34] to compute interaction energies between the ligand and individual amino acids defining the binding site along the MD simulations. Ligand-residue interaction energies are calculated as the sum of the (OPLS2005) van der Waals and electrostatic terms. This dynamic-interaction profile is saved as a time series in a comma-separated-values (csv) file

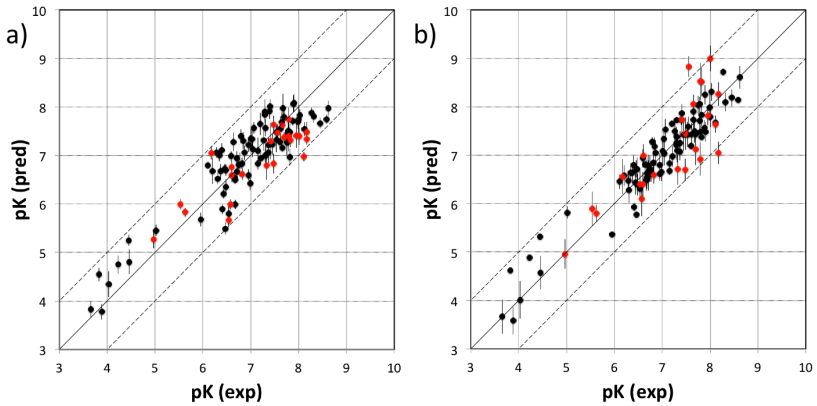
for subsequent download by the user or for interactive plotting and analysis using the *MD Client* front-end interface.

3. Results and Discussion

3.1. VirtualDesignLab

The current FimH Quasar model for the *VirtualDesignLab* was established based on structural and biological data of 108 mannose-based inhibitors (with IC_{50} values ranging from 220 μM to 2.4 nM) displaying diverse PK/PD profiles. Compound synthesis, biological assays [53] and model development were performed in-house, ensuring consistency of all results. Table 1 shows the structures, experimental and predicted affinities of compounds used for developing the model. The QSAR model based on a genetic algorithm converged at a cross-validated r^2 of 0.805 and yielded a predictive r^2 of 0.596 (Figure 6a). The only modest value of the predictive r^2 is a consequence of the relatively narrow range of test compound affinities (as some substances were necessary for the training set due to their structural uniqueness). The performance of the model is therefore better reflected by the individual predictions (23 out of 24 test substances within a factor of 10 from their experimental affinity). We further challenged the FimH model by using Y-scrambling and consensus scoring with the software Raptor (dual-shell 5D-QSAR; Figures 6b and S1) [54]. All tests, including the processing of additional external compounds, confirmed the predictive power of the mQSAR model-based framework.

Figure 6. Comparison between experimental (horizontal axis) and predicted pK values (vertical axis) for (**a**) the Quasar model and (**b**) the Raptor model. Black and red points represent compounds of the training and test set, respectively. Vertical bars indicate the estimated standard deviations of the prediction. Dashed lines are drawn at factors of 10 from the experimental value.



The *VirtualDesignLab* is aimed at predicting the binding affinity for a given compound within a factor of 10 from the experimental value. Currently, the affinity prediction for a single compound requires approximately one hour of CPU time—a good balance between accuracy and processing time. Special treatment may, however, be required for compounds retaining flexibility upon binding. In such

cases, improved entropy estimation (a method is currently in development at our institute) or a non-static approach, such as the one offered by the *MD Client* (cf. below), may be necessary. We would like to emphasize that the framework is independent of the FimH mQSAR model (presented here), as it only requires the generation and validation of a new QSAR model for any target protein of interest. This can be developed using by freely available software, e.g., Quasar [35].

3.2. MD Client

In the *MD Client*, the outcome of an MD simulation includes a set of frames extracted from the MD trajectory and a dynamic-interaction profile comprising per-residue interaction energies between ligand and protein for all time points. An interactive plot of computed interaction profiles is readily accessible from the *MD Client* interface (Figure 7). The plot created by the *matplotlib* python extension is cumulative, *i.e.*, it can incorporate dynamic profiles from several simulations in the same plot with automated coloring and legend generation. Comparing dynamic profiles of different simulations may provide valuable clues, for instance, about interaction modes of different ligands and/or key residues in ligand recognition and binding. We chose five structurally distinct FimH ligands (9, 17, 18, 28 and 37) to demonstrate the usefulness of dynamic-interaction profiles. Examination of their dynamic interaction energies with two key FimH residues (Gln133 and Phe1) indicates that these interactions are maintained throughout the entire simulation and that they don't significantly differ among different ligand classes (Figure 7a,b). These residues are typically involved in an extended hydrogen-bond network with the mannose moiety common to FimH binders. The profiles also show that the interaction with the N-terminal NH_3^+ moiety of Phe1 results in a considerably larger contribution to the binding enthalpy compared to Gln133. Automated docking of FimH ligands typically predicts a hydrogen bond from the 3-OH of the mannose moiety to the Asp140 residue to be thermodynamically favorable. Interestingly, this hydrogen bond does not seem to be kinetically stable, since it is broken within the first 0.5 ns and is never re-established throughout the entire simulation as can be observed in the profiles of all studied compounds (Figure 7c).

Figure 7. Dynamic per-residue interaction plots for five FimH ligands generated by the interactive plotting feature of *MD Client*; (**a**) Gln133, (**b**) Phe1, (**c**) Asp140 and (**d**) Tyr48. Vertical axis: protein–ligand interaction energies (kcal/mol); horizontal axis: molecular dynamics simulation time (ns). The colors mark the individual compounds shown above.

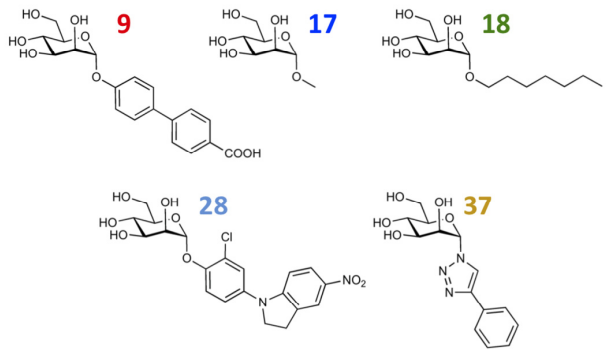
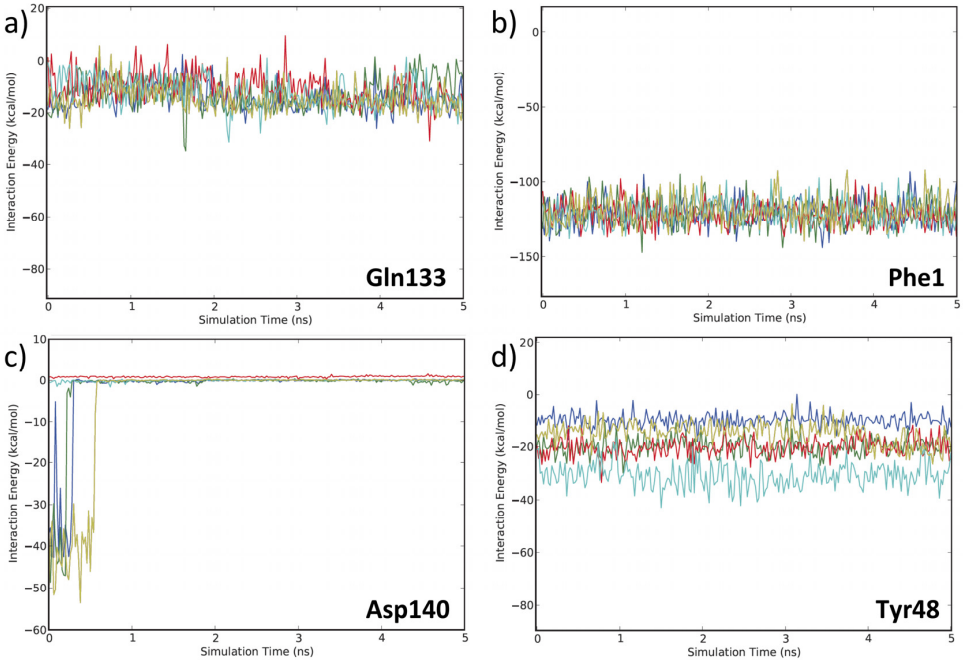
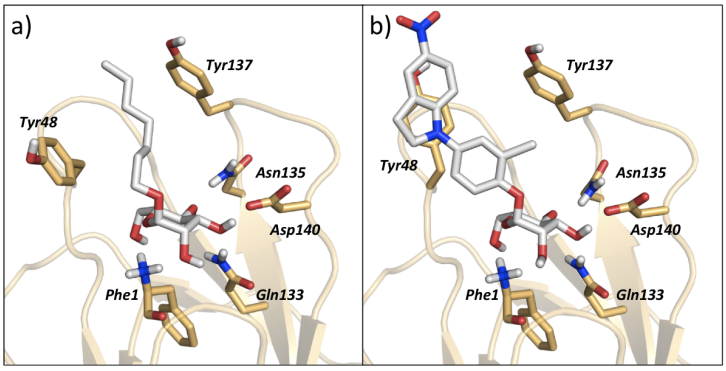


Figure 7. *Cont.*



The various classes of FimH ligands differ in their interaction patterns with the so-called *tyrosine gate* lined by Tyr48 and Tyr137 located at the entrance to the mannose-binding site (Figure 8) [23]. The interaction dynamics show that compound 28 exhibits the strongest interaction with Tyr48, which could be explained by its unique scaffold that allows a preferred interaction with the tyrosine side chain (Figure 7d). Compounds 9 and 18 display a favorable interaction with Tyr48, yet of lower magnitude than 28, which explains the superior affinity of the latter (see Table 1). Finally, two ligands seem to lack this favorable interaction with Tyr48; namely 17 and 37, which also coincides with their relatively lower affinities. This could be rationalized by the lack of a side chain capable of interacting with the Tyr48 in 17 and the inherent rigidity of 37 due to a lack of glycosidic oxygen in the linker between mannose and the aromatic aglycone.

Figure 8. Illustration of (**a**) *in* binding mode of 18 and (**b**) *out* binding mode of 28, within the tyrosine gate of FimH binding site. References to different binding modes of FimH ligands are given in the methods section.



4. Software Extension/Repurposing

The philosophy behind all our software is to allow for extendibility, as well as redirection, towards different targets of interest. In the following, we provide a brief overview of the corresponding requirements.

4.1. VirtualDesignLab

Assuming that a decent number of ligands with experimental affinity data are available for the given target, a three-dimensional protein-ligand structure that will serve for automated, flexible docking is required. These are usually obtained by means of crystallography or homology modeling and must typically be further refined (addition of hydrogen atoms, completion of missing amino-acid residues or their side chains, completion/generation of the solvent shell, energy minimization). These tasks can be accomplished through numerous, freely accessible computational tools. With the structure at hand, potential binding poses of all tested compounds need to be obtained. For this step, any flexible-docking software may be employed, including Alignator/Cheetah discussed in this article. The ensemble of potential binding modes can be compiled into a 4D data set to serve as input for the generation of the binding site surrogate. Though this task is best handled using the Quasar software, a QSAR model of different origin could also potentially be utilized. It should be noted, however, that even though structure preparation and pose generation are relatively simple tasks requiring no more than a few days of work, developing and validating a robust and reliable QSAR model is a complex and lengthy procedure. Also, given the vast diversity of the computational methods, personal communication with the authors of this article would likely be necessary in order to integrate a QSAR model with the *VirtualDesignLab* framework.

4.2. MD Client

The *MD Client* relies on the Desmond package at its back-end terminal to perform MD simulations. Desmond can be obtained free of charge for academics and non-commercial users. Once a working Desmond installation is available, the user needs to point the *MD Client* to where the back-end is located by providing the necessary SSH credentials. The *MD Client* can basically take it from there, since it has internal routines for identifying protein and ligand, job submission and monitoring, as well as calculating, organizing and plotting the interaction energy results.

5. Conclusions

Over the past three decades, much progress has been made in developing and validating innovative computational algorithms for common drug design-related tasks. In their perspective on the future of medicinal chemistry, Satyanarayanajois and Hill [55] stated that emerging medicinal chemists should additionally acquire “computational and cheminformatics acumen considerably greater than in years past”. In a related analysis, Ritchie and McLay [56] concluded that the goal of encouraging medicinal chemists to rely more on computational chemistry tools could be best achieved via specially designed tools that are “well-thought-out, suitable for their needs, able to generate useful, timely and valid

results". Similarly, we trust that adopting this strategy will ultimately maximize the benefit of state-of-the-art modeling technologies in the field of drug design and development.

To this end, we designed versatile single-click tools to assist medicinal chemists in performing two routine modeling tasks: the *VirtualDesignLab* for predicting binding mode and affinity of potential drug candidates and the *MD Client* for investigating dynamic behavior and energetics of ligand-protein complexes. Thanks to their modular design based mainly on self-developed algorithms, our tools allow easy modification, extension, as well as reorientation towards other targets and platforms of interest. Our group previously introduced the *OpenVirtualToxLab* for prediction of the toxic potential of drug candidates and made it freely available to academic organizations [57]. The two new tools introduced in this article, *VirtualDesignLab* and *MD Client*, can also be made available on request. Our future plans for *MD Client* include adding support for more molecular dynamics packages, as well as more analysis functionalities (for instance, surface area and entropy computations) to give more insight into ligand–protein interaction processes.

In closing, we wish to emphasize that our tools are not intended to, neither can they, replace the expert molecular modeler. In fact, their main purpose is to facilitate handling routine drug-design related tasks, thus leaving the more time-consuming detailed investigation only for interesting cases. Our vision is to place our interfacing technologies right on the medicinal chemists' workbench and keep all the complicated 'machinery' on a transparently maintained server. However, the simplicity of such tools, although tempting, should never cloud the medicinal chemists' judgment. On the contrary, medicinal chemists should always employ their expertise to question the results obtained from such tools.

Acknowledgments

Financial support from the Swiss National Foundation (grant No. 315230_122478) and the Department of Pharmaceutical Sciences, University of Basel, is gratefully acknowledged.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Bottegoni, G. Protein-ligand docking. *Front. Biosci.* **2012**, *17*, 2289–2306.
2. Kitchen, D.B.; Decornez, H.; Furr, J.R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat. Rev. Drug Discov.* **2004**, *3*, 935–949.
3. Nantasenamat, C.; Isarankura-Na-Ayudhya, C.; Naenna, T.; Prachayasittikul, V. A practical overview of Quantitative Structure-Activity Relationship. *EXCLI Journal* **2009**, *8*, 74–88.
4. Verma, J.; Khedkar, V.M.; Coutinho, E.C. 3D-QSAR in drug design—A review. *Curr. Top. Med. Chem.* **2010**, *10*, 95–115.
5. Karplus, M.; McCammon, J.A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646–652.

6. Cheatham, T.E.; Kollman, P.A. Molecular dynamics simulation of nucleic acids. *Annu. Rev. Phys. Chem.* **2000**, *51*, 435–471.
7. Alonso, H.; Bliznyuk, A.A.; Gready, J.E. Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.* **2006**, *26*, 531–568.
8. Rapaport, D.C. *The Art of Molecular Dynamics Simulation*. Cambridge University Press: Cambridge, UK, 2004; pp. 1–10.
9. Törnroth-Horsefield, S.; Wang, Y.; Hedfalk, K.; Johanson, U.; Karlsson, M.; Tajkhorshid, E.; Neutze, R.; Kjellbom, P. Structural mechanism of plant aquaporin gating. *Nature* **2006**, *439*, 688–694.
10. Beckstein, O.; Tai, K.; Sansom, M.S. Not ions alone: Barriers to ion permeation in nanopores and channels. *J. Am. Chem. Soc.* **2007**, *126*, 14694–14695.
11. Berendsen, H.J.C. Simulating the Physical World: Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics; Cambridge University Press: Cambridge, UK; 2007; p. 624.
12. Muegge, I. Synergies of virtual screening approaches. *Mini Rev. Med. Chem.* **2008**, *8*, 927–933.
13. Kirchmair, J.; Distinto, S.; Schuster, D.; Spitzer, G.; Langer, T.; Wolber, G. Enhancing drug discovery through in silico screening: Strategies to increase true positives retrieval rates. *Curr. Med. Chem.* **2008**, *15*, 2040–2053.
14. Schrödinger Suite. Available online: http://www.schrodinger.com (accessed on 12 December 2012).
15. Accelrys Discovery Studio. Available online: http://accelrys.com/products/discovery-studio (accessed on 12 December 2012).
16. SYBYL-X Suite. Available online: http://tripos.com (accessed on 12 December 2012).
17. Molecular Operating Environment. Available online: http://www.chemcomp.com (accessed on 12 December 2012).
18. Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K.; Olson, A.J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
19. Vedani, A.; Dobler, M. 5D-QSAR: The key for simulating induced fit? *J. Med. Chem.* **2002**, *45*, 2139–2149.
20. Bowers, K.J.; Chow, E.; Xu, H.; Dror, R.O.; Eastwood, M.P.; Gregersen, B.A.; Klepeis, J.L.; Kolossvary, I.; Moraes, M.A.; Sacerdoti, F.D.; *et al.* Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)*, Tampa, FL, USA, 11–17 November 2006.
21. Irwin, J.J.; Shoichet, B.K.; Mysinger, M.M.; Huang, N.; Colizzi, F.; Wassam, P.; Cao, Y. Automated docking screens: A feasibility study. *J. Med. Chem.* **2009**, *52*, 5712–5720.
22. Klein, T.; Abgottspon, D.; Wittwer, M.; Rabbani, S.; Herold, J.; Jiang, X.; Kleeb, S.; Luthi, C.; Scharenberg M.; Bezencon, J.; *et al*. FimH Antagonists for the oral treatment of urinary tract infections: From design and synthesis to *in vitro* and *in vivo* evaluation. *J. Med. Chem.* **2010**, *53*, 8627–8641.
23. Schwardt, O.; Rabbani, S.; Hartmann, M.; Abgottspon, D.; Wittwer, M.; Kleeb, S.; Zalewski, A.; Smieško, M.; Cutting, B.; Ernst, B. Design, synthesis and biological evaluation of mannosyl triazoles as FimH antagonists. *Bioorg. Med. Chem.* **2011**, *19*, 6454–6473.

24. Jiang, X.; Abgottspon, D.; Kleeb, S.; Rabbani, S.; Scharenberg, M.; Wittwer, M.; Haug, M.; Schwardt, O.; Ernst, B. Antiadhesion therapy for urinary tract infections—A balanced PK/PD profile proved to be key for success. *J. Med. Chem.* **2012**, *55*, 4700–4713.
25. QikProp, version 3.4; Schrödinger, LLC: New York, NY, USA, 2011.
26. Vedani, A.; Dobler, M.; Smieško, M. VirtualToxLab—A platform for estimating the toxic potential of drugs, chemicals and natural products. *Toxicol. Appl. Pharmacol.* **2012**, *261*, 142–153.
27. VirtualToxLab Online Documentation. Available online: http://www.biograf.ch/downloads/VirtualToxLab.pdf (accessed on 12 December 2012).
28. MacroModel, version 9.9; Schrödinger, LLC: New York, NY, USA, 2011.
29. Storer, J.W.; Giesen, D.J.; Cramer, C.J.; Truhlar, D.G. Class IV charge models: A new semiempirical approach in quantum chemistry. *J. Comput. Aided Mol. Des.* **1995**, *9*, 87–110.
30. ConfGen, version 2.1; Schrödinger, LLC: New York, NY, USA, 2009.
31. Smieško, M. University of Basel, Basel, Switzerland. Personal communication, 2012.
32. Rossato, G.; Ernst, B.; Smiesko, M.; Spreafico, M.; Vedani, A. Probing small-molecule binding to cytochrome P450 2D6 and 2C9: An *in silico* protocol for generating toxicity alerts. *ChemMedChem* **2010**, *5*, 2088–2101.
33. Han, Z.; Pinkner, J.S.; Ford, B.; Obermann, R.; Nolan, W.; Wildman, S.A.; Hobbs, D.; Ellenberger, T.; Cusumano, C.K.; Hultgren, S.J.; *et al*. Structure-based drug design and optimization of mannoside bacterial FimH antagonists. *J. Med. Chem.* **2010**, *53*, 4779–4792.
34. Maestro, version 9.2; Schrödinger, LLC: New York, NY, USA, 2011.
35. Vedani, A.; Dobler, M.; Lill, M.A. Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. *J. Med. Chem.* **2005**, *48*, 3700–3703.
36. Vedani, A.; Huhta, D.W. A new force field for modeling metalloproteins. *J. Amer. Chem. Soc.* **1990**, *112*, 4759–4767.
37. Rojewska, D.; Elber, R. Molecular dynamics study of secondary structure motions in proteins: application to myohemerythrin. *Proteins* **1990**, *7*, 265–279.
38. Wan, S.; Flower, D.R.; Coveney, P.V. Toward an atomistic understanding of the immune synapse: Large-scale molecular dynamics simulation of a membrane-embedded TCR-pMHC-CD4 complex. *Mol. Immunol.* **2008**, *45*, 1221–1230.
39. Knapp, B.; Omasits, U.; Bohle, B.; Maillere, B.; Ebner, C.; Schreiner, W.; Jahn-Schmid, B. 3-Layer-based analysis of peptide-MHC interaction: *In silico* prediction, peptide binding affinity and T cell activation in a relevant allergen-specific model. *Mol. Immunol*. **2009**, *46*, 1839–1844.
40. Coll, E.P.; Kandt, C.; Bird, D.A.; Samuels, A.L.; Tieleman, D.P. The distribution and conformation of very long-chain plant wax components in a lipid bilayer. *J. Phys. Chem. B* **2007**, *111*, 8702–8704.
41. Kandt, C.; Ash, W.L.; Tieleman, D.P. Setting up and running molecular dynamics simulations of membrane proteins. *Methods* **2007**, *41*, 475–488.
42. Heller, H.; Schaefer, M.; Schulten, K. Molecular dynamics simulation of a bilayer of 200 lipids in the gel and in the liquid crystal phase. *J. Phys. Chem.* **1993**, *97*, 8343–8360.

43. Miller, J.L.; Kollman, P.A. Theoretical studies of an exceptionally stable RNA tetraloop: Observation of convergence from an incorrect NMR structure to the correct one using unrestrained molecular dynamics. *J. Mol. Biol.* **1997**, *270*, 436–450.
44. Luo, J.; Bruice, T.C. Nanosecond molecular dynamics of hybrid triplex and duplex of polycation deoxyribonucleic guanidine strands with a complimentary DNA strand. *J. Amer. Chem. Soc.* **1998**, *120*, 1115–1123.
45. Case, D.A.; Cheatham, T.E., III.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
46. Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.
47. Christen, M.; Hünenberger, P.H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D.P.; Heinz, T.N.; Kastenholz, M.A.; Kräutler, V.; Oostenbrink, C.; *et al*. The GROMOS software for biomolecular simulation: GROMOS05. *J. Comp. Chem.* **2005**, *26*, 1719–1751.
48. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
49. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
50. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.
51. Desmond Molecular Dynamics System, version 3.1; D. E. Shaw Research: New York, NY, USA, 2012.
52. Maestro-Desmond Interoperability Tools, version 3.1; Schrödinger: New York, NY, USA, 2012.
53. Rabbani, S.; Jiang, X; Schwardt, O.; Ernst, B. Expression of the carbohydrate recognition domain of FimH and development of a competitive binding assay. *Anal. Biochem.* **2010**, *407*, 188–195.
54. Lill, M.A.; Vedani, A.; Dobler, M. Raptor: Combining dual-shell representation, induced-fit simulation and hydrophobicity scoring in receptor modeling: Application toward the simulation of structurally diverse ligand sets. *J. Med. Chem.* **2004**, *47*, 6174–6186.
55. Satyanarayanajois, S.D.; Hill, R.A. Medicinal chemistry for 2020. *Future Med. Chem.* **2011**, *3*, 1765–1786.
56. Ritchie, T.J.; McLay, I.M. Should medicinal chemists do molecular modelling? *Drug Discov. Today* **2012**, *17*, 534–537.
57. Open VirtualToxLab. Available online: http://www.virtualtoxlab.org (accessed on 12 December 2012).

7.6 *BioorgMedChem.* 2011; *19(21)*: 6454-73.

Bioorganic & Medicinal Chemistry 19 (2011) 6454–6473

Contents lists available at SciVerse ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

Design, synthesis and biological evaluation of mannosyl triazoles as FimH antagonists

Oliver Schwardt, Said Rabbani, Margrit Hartmann, Daniela Abgottspon, Matthias Wittwer, Simon Kleeb, Adam Zalewski, Martin Smieško, Brian Cutting, Beat Ernst *

Institute of Molecular Pharmacy, Pharmacenter, University of Basel, Klingelbergstrasse 50, CH-4056 Basel, Switzerland

ARTICLE INFO

Article history:
Received 22 April 2011
Revised 24 August 2011
Accepted 25 August 2011
Available online 31 August 2011

Keywords:
Uropathogenic *Escherichia coli*
Urinary tract infections
Bacterial adhesin FimH
FimH antagonists
Competitive binding assay
Aggregometry assay
NMR spectroscopy

ABSTRACT

Urinary tract infection (UTI) caused by uropathogenic *Escherichia coli* (UPEC) is one of the most prevalent infectious diseases. Particularly affected are women, who have a 40–50% risk to experience at least one symptomatic UTI episode at some time during their life. In the initial step of the infection, the lectin FimH, located at the tip of bacterial pili, interacts with the high-mannosylated uroplakin Ia glycoprotein on the urinary bladder mucosa. This interaction is critical for the ability of UPEC to colonize and invade the bladder epithelium. X-ray structures of FimH co-crystallized with two different ligands, the physiological binding epitope oligomannose-3 and the antagonist biphenyl α-D-mannoside **4a** revealed different binding modes, an *in-docking-mode* and an *out-docking-mode*, respectively. To accomplish the *in-docking-mode*, that is the docking mode where the ligand is hosted by the so-called tyrosine gate, FimH antagonists with increased flexibility were designed and synthesized. All derivatives **5**–**8** showed nanomolar affinities, but only one representative, the 4-pyridiyl derivative **5j**, was as potent as the reference compound *n*-heptyl α-D-mannoside (**1b**). Furthermore, a loss of affinity was observed for *C*-glycosides and derivatives where the triazole aglycone is directly *N*-linked to the anomeric center. A conformational analysis by NMR revealed that the triazolyl-methyl-*C*-mannosides **8** adopt an unusual $^{1}C_{4}$ chair conformation, explaining the comparably lower affinity of these compounds. Furthermore, to address the drug-likeness of this new class of FimH antagonists, selected pharmacokinetic parameters, which are critical for oral bioavailability (lipophilicity, solubility, and membrane permeation), were determined.

1. Introduction

Urinary tract infections (UTIs) are among the most common infections, affecting millions of people each year. Although UTIs rarely cause severe diseases such as pyelonephritis or urosepsis, they are associated with extensive morbidity and generate considerable medical expenses.[1] Uropathogenic *Escherichia coli* (UPEC) are the primary cause of UTIs accounting for 70–95% of the reported cases. Particularly affected are women, who have a 40–50% risk to suffer from a symptomatic UTI episode at some time during their life.[2,3] Symptomatic UTIs require antimicrobial treatment, resulting in selection and development of bacterial resistance. Consequently, treatment of consecutive infections becomes increasingly difficult. Especially patients with diabetes, urinary tract anomaly, paraplegia and those with permanent urinary catheter experience repeated UTIs with resistant strains. Therefore, a new approach for the treatment and prevention of UTI with non-antibiotic and orally applicable therapeutics with a low potential for resistance would have a great impact on patient care, public healthcare, and medical expenses.

UPEC express a number of well-studied virulence factors for successful colonization of and survival within the host.[1,4,5] One important virulence factor, the mannose-specific FimH adhesin, is located at the tip of bacterial type 1 pili.[6] Type 1 pili are the most prevalent fimbriae encoded by UPEC, consisting of the four subunits FimA, FimF, FimG and FimH. The FimH lectin enables UPEC to attach to high-mannosylated uroplakin Ia glycoproteins on the urinary bladder mucosa, thus enabling adherence and invasion of host cells and at the same time preventing the rapid clearance of

Abbreviations: ABTS, 2,2′-azino-di-(3-ethylbenzthiazoline-6-sulfonic acid); AUC, area under the curve; BSA, bovine serum albumin; CRD, carbohydrate recognition domain; D, distribution coefficient; DCM, dichloromethane; DMSO, dimethyl sulfoxide; GIT, gastrointestinal tract; GPE, guinea pig erythrocytes; HEPES, 4-(2-hydroxyethyl)-piperazine-1-ethanesulfonic acid; IC_{50}, half maximal inhibitory concentration; iv, intravenous; Man, D-mannose; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser enhancement spectroscopy; PAA, polyacrylamide; PAMPA, parallel artificial membrane permeation assay; P_{app}, apparent permeability; P_e, effective permeation; po, peroral; rIC_{50}, relative inhibitory concentration; SAR, structure–activity relationship; THF, tetrahydrofurane; TM-PAA, Manα(1-3)-[Manα(1-6)]-Manβ(1-4)-Glc*N*Acβ(1-4)-Glc*N*Acβ-PAA; UPEC, uropathogenic *E. coli*; UTI, urinary tract infection.

* Corresponding author. Tel.: +41 61 2 67 15 51; fax: +41 61 2 67 15 52.
E-mail address: beat.ernst@unibas.ch (B. Ernst).

0968-0896/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2011.08.057

E. coli from the UTI by the bulk flow of urine.[1,7] As a part of the FimH subunit, a carbohydrate-recognition domain (CRD) is responsible for bacterial interactions with the host cells within the urinary tract.[7] The crystal structure of methyl α-D-mannopyranoside bound to the FimH-CRD was solved[8] and the structures of the corresponding complexes with *n*-butyl α-D-mannopyranoside,[9] Manα(1-3)-[Manα(1-6)]-Manβ(1-4)-Glc*N*Acβ-(1-4)Glc*N*Ac (oligomannose-3)[10] and biphenyl α-D-mannopyranoside[11] recently became available.

Previous studies showed that colonization and subsequent *E. coli* infection of the human urothelium can be prevented by vaccination with FimH adhesin.[12,13] Furthermore, adherence and invasion of host cells by *E. coli* can also be inhibited by oligomannosides representing the glycosylation of uroplakin **1a**.[14] For some α-D-mannosides it was shown that they prevent type 1 pili mediated adhesion, that is, they do not act by killing or arresting the growth of the pathogen as antibiotics do. Therefore, the spread of strains resistant to such agents are expected to be significantly delayed as compared to that of strains resistant to antibiotics.[15] In addition, environmental contamination is less problematic compared to antibiotics.[15a]

More than two decades ago, various oligomannosides and aromatic α-D-mannosides that antagonize type 1 fimbriae-mediated bacterial adhesion were identified.[15,16] However, for these mannosides only weak interactions in the milli- to micromolar range were observed. To improve their affinity, the multivalent presentation of the α-mannoside epitope,[17] and the rational design of ligands guided by structural information were explored.[9–11] Recently, various reports on high affinity monovalent FimH antagonists were published.[11,18,19]

The CRD of the FimH protein consists of amino acids with hydrophilic side chains and can therefore establish a perfect network of hydrogen bonds with the hydroxyl groups at the 2-, 3-, 4- and 6-positions of D-mannose. The entrance to this mannose-binding pocket, the so-called 'tyrosine gate', is shaped by two tyrosines (Tyr48 and Tyr137), and one isoleucine (Ile52) which support hydrophobic contacts.[20] Generally, long chain alkyl and aryl mannosides (for selected examples see Fig. 1) displayed the highest affinities.[8,9,11,16–21]

Recently, we reported the synthesis, the critical pharmacokinetic properties and affinity data of low molecular weight α-D-mannosides with the ability to block the FimH-mediated bacterial adhesion in a mouse infection model.[19] The orally available, nanomolar FimH antagonist **4b** (Fig. 1) exhibited the potential to reduce the colony forming units (CFU) in the urine and in the bladder by two and four orders of magnitude, respectively, demonstrating the therapeutic potential of this new class of anti-infectives for the effective treatment of urinary tract infections.

However, a potential drawback of FimH antagonists with aglycons consisting of biphenyls directly linked to the carbohydrate moiety is their limited conformational flexibility, which could hamper an optimal fit with the tyrosine gate.[11] To increase the conformational flexibility, the spacers between the mannose moiety and the first aromatic ring of the biphenyl moiety in **i** (Fig. 2) as well as between the aromatic rings was extended. Furthermore, the rotational barrier of the biphenyl[25] was reduced by replacing one of the rings by a triazole (for the torsion profile see Fig. 2). Overall, these modifications should lead to a reduction of the conformational restraints and therefore an optimized spatial arrangement of the aglycone in the tyrosine gate.

Oligomannose-3 is present on the high-mannosylated uroplakin Ia located on urothelial cells and is supposed to interact with UPEC. The crystal structure of the FimH-CRD[10] complexed with oligomannose-3 (PDB code 2VCO, Fig. 3A) clearly shows the important role of the tyrosine gate hosting this physiological ligand in the so-called *in-docking-mode*. Interestingly, for **4a** complexed with FimH-CRD a different binding mode outside of the tyrosine gate was reported (*out-docking-mode*, see Fig. 3B).[11] In analogy to oligomannose-3, docking of triazole derivative **5b** to the crystal structure of the FimH lectin domain (PDB code 3MCY)[11] led - as a result of the increased flexibility of the aglycone - to the *in-docking-mode*. Thus, in contrast to the biphenyl aglycone present in **4a**, the phenyl-triazole **5b** is expected to be hosted by the tyrosine gate. The three-dimensional structure **5b** was generated using Glide 5.5[26] and the kinetic stability of the protein–ligand complex was then assessed with a 2 ns molecular-dynamics simulation using Desmond.[27]

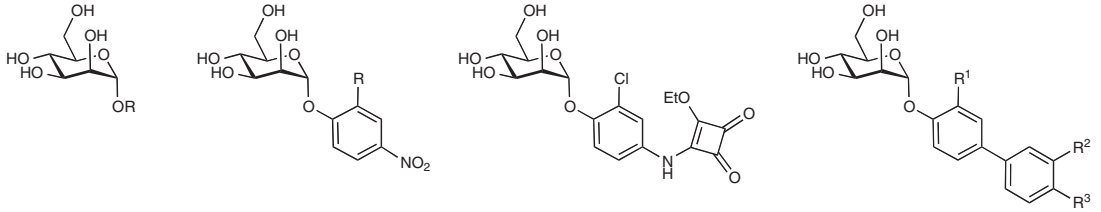
A comparison of the docking modes of oligomannose-3, **4a** and **5b** reveals that the interaction of the mannose moiety is highly conserved for all three compounds. However, in contrast to oligomannose-3 and **5b**, the biphenyl moiety in **4a** is not able to reach the tyrosine gate due to its rigid structure. Instead, a π-π-stacking interaction of the second aromatic ring of the biphenyl aglycone with Tyr48 outside of the tyrosine gate[11] (*out-docking-mode*, Fig. 3B) is achieved by induced fit, that is, a substantial move of Tyr48. In addition, a further stabilization of the protein–ligand complex by a hydrogen bond between the ester in the *meta*-position of **4a** and the side-chain of Arg98 was assumed.[11]

Based on these evidences, a library of derivatives according to the criteria summarized in Figure 2 was designed. Here, we describe synthesis, biological evaluation, and determination of pharmacokinetic parameters of triazole derivatives.

2. Results and discussion

2.1. Synthesis of triazolyl-methyl and -ethyl α-D-mannopyranosides

In a first approach, the phenyl ring adjacent to the anomeric center (see Fig. 2) was replaced by a triazolyl-methyl moiety to increase the conformational flexibility. To avoid solubility problems as well as to take advantage of additional polar interactions, for



1a (R: Me)
1b (R: *n*-heptyl)[9]

2a (R: H)[18a,b]
2b (R: Cl)[18c]

3[18c]

4a (R^1: H, R^2: CO_2Me, R^3: H)[11]
4b (R^1: Cl, R^2: H, R^3: CO_2Me)[19]
4c (R^1: Cl, R^2: H, R^3: CO_2Na)[19]

Figure 1. Known alkyl (**1**) and aryl (**2**–**4**) α-D-mannosides exhibiting micro- to nanomolar affinities.

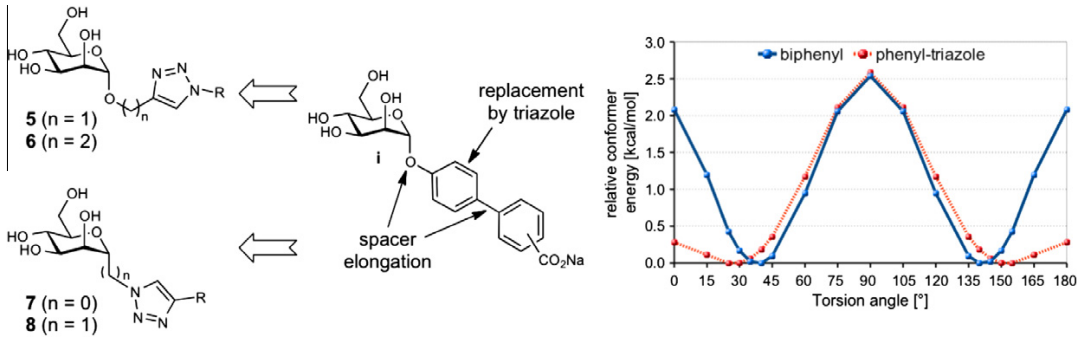


Figure 2. Design of FimH antagonists with aglycons of increased flexibility. Spacer elongations and replacement of one phenyl ring by a triazole should reduce the conformational restraints and lead to an improved fit in the tyrosine gate. The torsion profiles for biphenyl and 1-phenyl-1,2,3-triazole were calculated at the B3LYP level of theory[22,23] with 6-31G(d,p) basis set in the gas phase using Gaussian 03.[24]

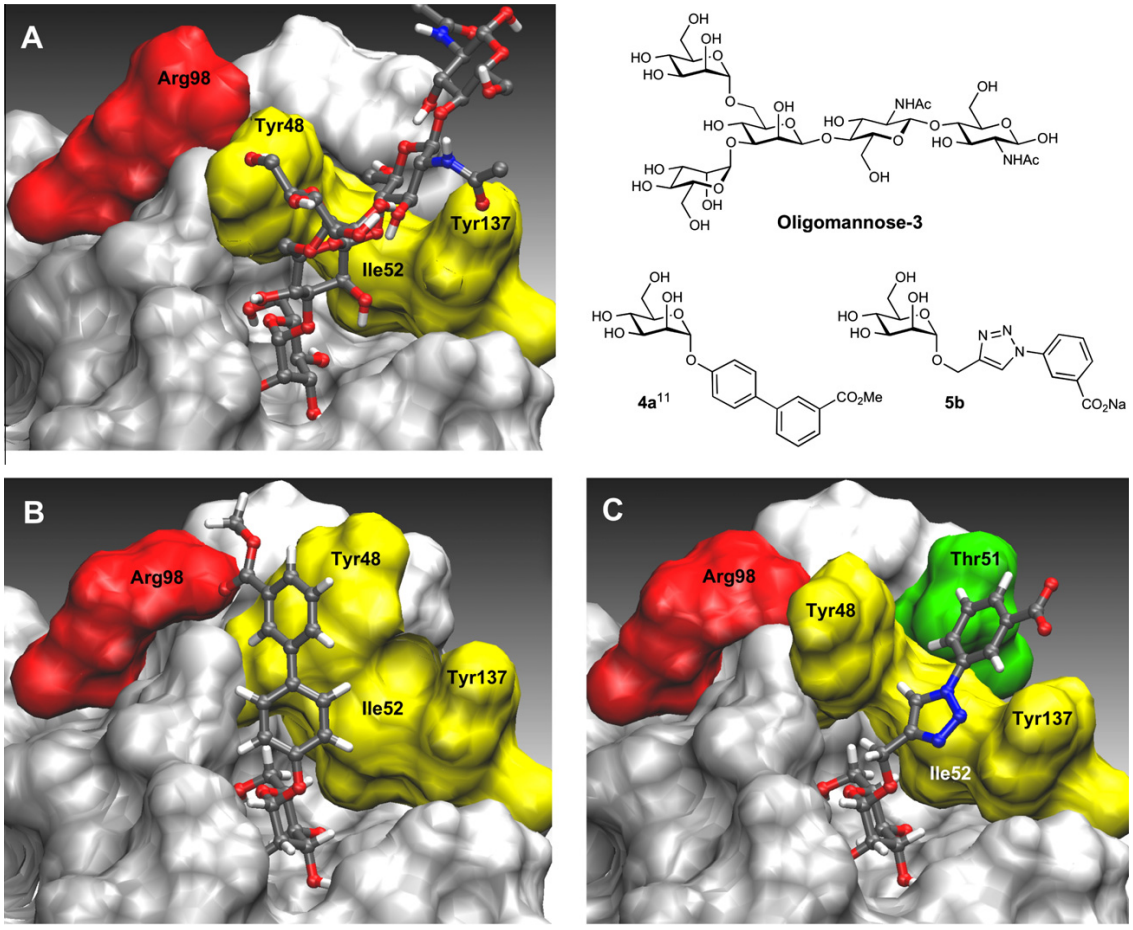
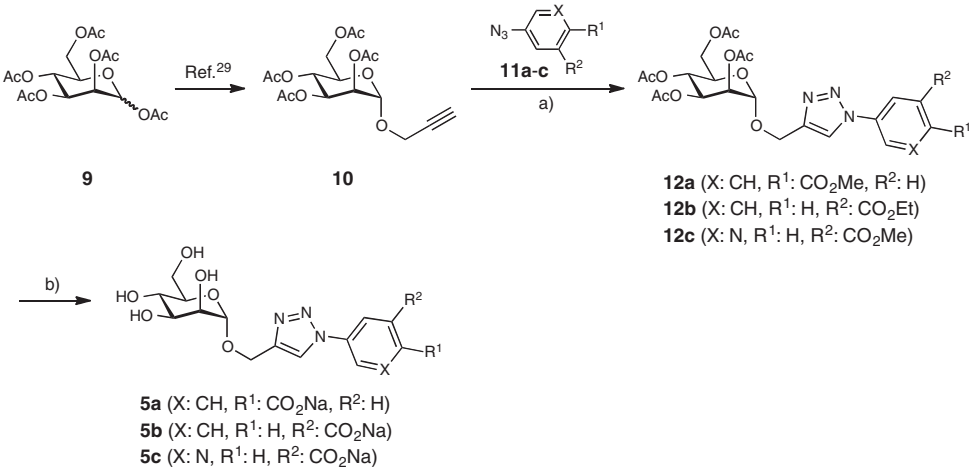


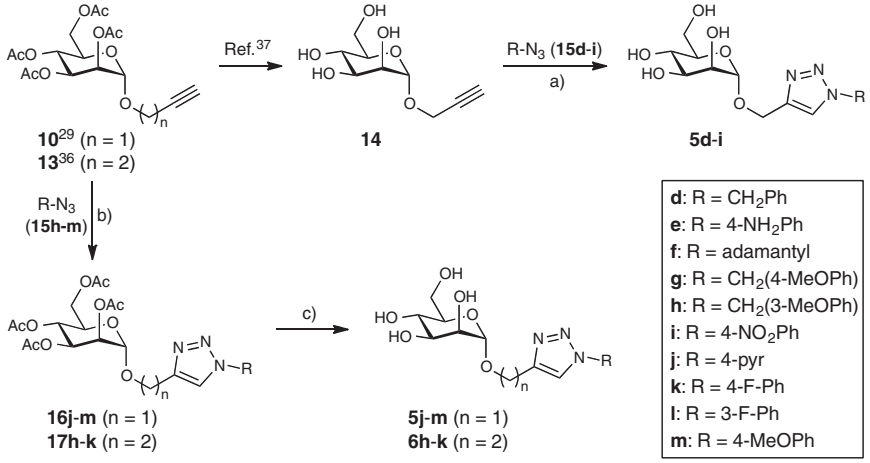
Figure 3. (A) & (B) Crystal structures of oligomannose-3 (A, PDB code 2VCO)[10] and biphenyl **4a** (B, PDB code 3MCY)[11] bound to the FimH-CRD. C) Automated docking of triazole **5b** into the lectin domain of FimH (PDB code 3MCY).[11] The images have been generated using VMD.[28] The ligands are depicted colored by atom (C: dark grey, H: white, O: red, N: blue); the tyrosine gate (residues Tyr48, Tyr137 and Ile52) is shown in yellow, residue Thr51 in green and residue Arg98 in red. While **4a** binds in the *out-docking-mode*, compound **5b**, like oligomannose-3, is inserted into the tyrosine gate (*in-docking-mode*).

example, H-bonds with the hydroxyl-groups of Thr51 or Tyr137 (Fig. 3C), the second aromatic ring was substituted with a carboxylate in *para*- or *meta*-position (→**5a**–**c**, Scheme 1).

For the synthesis of mannosyl triazoles **5a**–**c**, alkyne **10**[29] readily available from peracetylated D-mannose (**9**) was reacted with the known aryl azides **11a**,[30] **11b**,[31] and **11c**[32] in a copper(I)-catalyzed Huisgen 1,3-dipolar cycloaddition[33,34] using *tert*-butanol/water/THF (1:1:1) as solvent.[35] The saponification of the *anti*-substituted triazoles **12a**–**c** yielded the test compounds **5a**–**c** (Table 1).



Scheme 1. Reagents and conditions: (a) $CuSO_4{\cdot}5H_2O$, Na-ascorbate, *t*-BuOH/H_2O/THF (1:1:1), rt, 24 h, 73–97%; (b) (i) NaOMe, MeOH, rt, 3 h; (ii) 1 M NaOH, H_2O/dioxane (1:1), 16 h, 78–91%.



Scheme 2. Reagents and conditions: (a) $CuSO_4{\cdot}5H_2O$, Na-ascorbate, *t*-BuOH/H_2O (1:1), rt, 24 h, (**5d–i**: 27–77%); (b) $CuSO_4{\cdot}5H_2O$, Na-ascorbate, *t*-BuOH/H_2O/THF (1:1:1), rt, 24 h (**16j–m**: 85–94%, **17h–k**: 83–96%); (c) NaOMe, MeOH, rt, 2–6 h, (**5j–m**: 75–85%, **6h–k**: 73–90%).

In a second approach, the terminal aromatic ring was replaced by various substituents like (hetero)aryl, benzyl, and adamantyl groups (→**5d–i** & **5j–m**). Furthermore, in compounds **6h–k** the spacer between the carbohydrate moiety and the triazole ring was elongated from methyl to ethyl allowing for a higher conformational flexibility (Scheme 2).

The mannosyl triazoles **5d–m** and **6h–k** were obtained by reacting the known mannosyl alkynes **10**,[29] **13**[36] and **14**[37] with the azides **15d–m**. Whereas the azides **15d–f** are commercially available, **15g**,[38] **15h**,[39] **15i**,[40] **15j–l**,[41] and **15m**[40] were obtained by known procedures.

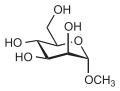
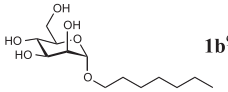
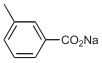
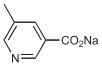
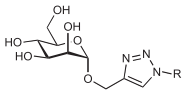
The cycloaddition of alkyne **14** and azides **15d–i** under Cu(I)-catalyzed click conditions[33,34] yielded directly the *anti*-substituted triazoles **5d–i** in 27–77% (Table 1). However, due to the cumbersome purification of the unprotected mannosyl triazoles, test compounds **5j–m** were obtained by an alternative sequence starting from the protected alkyne **10** and azides **15j–m** followed by saponification of the intermediates **16j–m**. The analogous cycloaddition of butinyl mannoside **13** with azides **15h–k** yielded the protected triazoles **17h–k** in 83–96%. Final deacetylation under Zemplén conditions gave the test compounds **6h–k**, which contain a linker extended by an additional carbon between mannose and aglycone (Table 1).

2.2. Synthesis of FimH antagonists modified at the anomeric center

To avoid the low metabolic stability of *O*-mannosides like compounds **5** and **6** due to potential cleavage by mannosidases, the corresponding *N*-linked mannosyl triazoles **7** and *C*-mannosides **8** were prepared (Scheme 3). Mannosyl azide **18** was obtained according to published procedures.[42] The Cu(I)-catalyzed click reaction of **18** with the commercially available acetylenes **19n–s** gave exclusively the anomerically pure *anti*-substituted α-D-mannosyl-triazoles **20n–s** in 84–98% yield and after deacetylation the test compounds **7n–s** (Table 1).

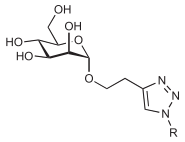
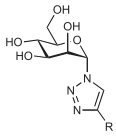
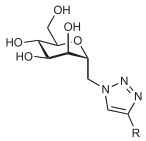
Finally, the synthesis of triazolyl-methyl-*C*-mannosides **8n–s** (Scheme 3) started from mannosyl cyanide **21**, which was obtained

Table 1
Pharmacodynamic and pharmacokinetic parameters of mannosylated triazoles **5**–**8**

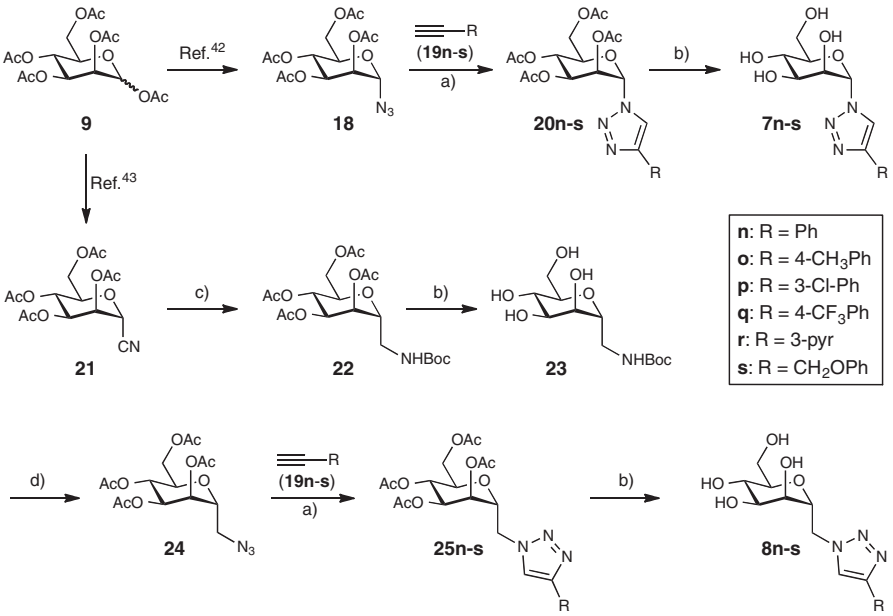
Entry	Ligand	R	Competitive binding assay		$logD_{7.4}$	PAMPA $logP_e$	Solubility
			IC_{50} [μM]	rIC_{50}		[$log10^{-6}$ cm/s]/%Mm	[μg/mL]
1		**1a**[45]	1.9	29	n.d.	n.d./n.d.	n.d.
2		**1b**[9]	0.0656	**1.0**	1.65	-4.89/21.0	> 3000
3		**5a** 	0.207	3.2	n.d.	n.d./n.d.	> 3000
4		**5b** 	0.296	4.5	n.d.	n.d./n.d.	> 3000
5		**5c** 	0.169	2.6	n.d.	n.d./n.d.	> 3000
6		**5d** 	0.667	10.2	-0.33	-8.60/3.0	> 3000
7		**5e** 	0.420	6.4	0.20	-10/4.0	> 2000
8		**5f** 	0.135	2.1	0.07	-7.42/9.8	> 3000
9		**5g** 	0.511	7.8	-0.40	-8.21/3.1	> 3000
10		**5h** 	0.452	6.9	-0.38	-8.18/4.0	> 3000
11		**5i** 	0.397	6.1	-0.16	-9.40/4.9	2600
12		**5j** 	0.070	1.1	0.24	-8.70/9.2	> 3000
13		**5k** 	0.778	11.9	-0.15	-10/4.6	> 3000
14		**5l** 	0.348	5.3	0.18	-8.80/5.4	> 3000
15		**5m** 	0.161	2.5	-0.05	-8.90/7.9	2600

(*continued on next page*)

Table 1. (*continued*)

Entry	Ligand	R	Competitive binding assay		$logD_{7.4}$	PAMPA $logP_e$ [$log10^{-6}$ cm/s]/%Mm	Solubility [μg/mL]
			IC_{50} [μM]	rIC_{50}			
16		**6h**	0.229	3.5	< -1	-8.70/11.8	> 3000
17		**6i**	0.112	1.7	-0.30	-9.20/2.9	> 3000
18		**6j**	0.153	2.3	< -1.5	-9.10/4.3	> 3000
19		**6k**	0.196	3.0	-0.21	-10/4.0	> 3000
20		**7n**	0.250	3.8	0.21	-8.6/6.4	> 3000
21		**7o**	0.248	3.8	0.91	-8.1/11.5	445
22		**7p**	0.331	5.1	1.22	-7.8/13.2	705
23		**7q**	0.144	2.2	1.45	-7.7/17.2	159
24		**7r**	0.216	3.3	-1.04	-10/10.3	1378
25		**7s**	0.493	7.5	-0.18	-9.0/5.5	> 3000
26		**8n**	0.560	8.5	-0.36	-9.1/7.8	> 3000
27		**8o**	0.565	8.6	0.23	-9.3/6.3	1489
28		**8p**	0.639	9.7	0.68	-9.7/n.p.	> 3000
29		**8q**	0.194	3.0	1.07	-9.3/n.p.	525
30		**8r**	0.333	5.1	< -1.5	-10/n.p.	1877
31		**8s**	0.327	5.0	-0.83	-10/n.p.	n.d.

The IC_{50}s were determined with a cell-free competitive binding assay.[45] Relative IC_{50}s (rIC_{50}) were calculated by dividing the IC_{50} of the substance of interest by the IC_{50} of the reference compound **1b** (entry 2). Passive permeation through an artificial membrane and retention therein was determined by PAMPA (parallel artificial membrane permeation assay).[50] Distribution coefficients ($logD_{7.4}$ values) were measured by a miniaturized shake flask procedure.[51] Thermodynamic solubility was measured by an equilibrium shake flask approach.[52] P_e effective permeation; n.p. not permeable; n.d. not determined.



Scheme 3. Reagents and conditions: (a) $CuSO_4{\cdot}5H_2O$, Na-ascorbate, *t*-BuOH/H_2O/THF (1:1:1), rt, 1–2 d (**20n–s**: 84–98%, **25n–s**: 93–98%); (b) NaOMe, MeOH, rt, 3–6 h, (**7n–s**: 65–92%, **8n–s**: 83–87%, **23**: 95%); (c) H_2 (4 bar), cat. Pd/C, Boc_2O, EtOAc, 1 d (72%); d) (i) concd HCl, dioxane/H_2O (2:1), 4 h; (ii) TfN_3, $NaHCO_3$, cat. $CuSO_4{\cdot}5H_2O$, PhMe/H_2O/MeOH, rt, 20 h; (iii) Ac_2O, pyridine, rt, 4 h (81%).

from **9** as reported earlier.[43] Catalytic hydrogenation in the presence of Boc_2O (→**22**) followed by deacetylation led to the Boc-protected amine **23**. Cleavage of the Boc-group, amine-azide exchange[44] and subsequent re-acetylation yielded azide **24** in 81% over three steps. The cycloaddition of **24** and the acetylenes **19n–s** gave the *anti*-substituted triazoles **25n–s** in excellent yields. Final deprotection afforded the test compounds **8n–s** (Table 1).

2.3. Biological evaluation

For an initial biological in vitro characterization, a cell-free competitive binding assay[45] and later on, a cell-based aggregation assay[46] were applied. Whereas for the cell-free competitive binding assay only the CRD of the pili was expressed, the complete pili are present in the cell-based assay format. Furthermore, both formats are competitive assays, that is, the analyzed antagonists compete with mannosides for the binding site. In the cell-free competitive binding assay, the competitors are polymer-bound trimannosides, whereas in the aggregation assay the antagonist competes with more potent oligo- and polysaccharide chains[14] present on the surface of erythrocytes.[47] The interaction is further complicated by the existence of a high- and a low-affinity state of the CRD of FimH. Aprikian et al. experimentally demonstrated that in full-length fimbriae the pilin domain stabilizes the CRD domain in the low-affinity state, whereas the CRD domain alone adopts the high-affinity state.[48] Furthermore, it was recently shown that shear stress can induce a conformational switch (twist in the β-sandwich fold of the CRD domain) resulting in improved affinity.[49] Therefore, different affinities are expected when - as in the cell-based aggregometry assay - full-length fimbriae are present, when compared to the CRD domain alone.

2.4. Cell-free competitive binding assay

The cell-free inhibition assay is based on the interaction of a biotinylated polyacrylamide glycopolymer with the carbohydrate recognition domain (CRD) of FimH as previously reported.[45] A soluble recombinant protein consisting of the FimH-CRD (amino acid residues 1–156), a C-terminal thrombin cleavage site and a 6His-tag (FimH-CRD-Th-6His) was expressed in *E. coli* strain HM125 and purified by affinity chromatography on a Ni-NTA column. For the determination of IC_{50} values, a microtiter plate coated with FimH-CRD-Th-6His was incubated with biotinylated Manα(1-3)-[Manα(1-6)]-Manβ(1-4)-Glc*N*Acβ(1-4)-Glc*N*Acβ-polyacrylamide (TM-PAA) polymer conjugated to streptavidin-horseradish peroxidase and the FimH antagonist in fourfold serial dilution (Fig. 4). The assay was performed in duplicates and repeated twice for each compound. To ensure comparability of different antagonists, the reference compound *n*-heptyl α-D-mannopyranoside (**1b**)[9,46] was tested in parallel on each individual microtiter plate. The affinities are reported relative to **1b** as rIC_{50} in Table 1. The relative IC_{50} (rIC_{50}) is the ratio of the IC_{50} of the test compound to the IC_{50} of **1b** (entry 2).

Interestingly, all antagonists in Table 1 except methyl α-D-mannoside (**1a**) exhibit nanomolar affinities. When compared to **1a**, an up to 30-fold improvement was obtained. In the first series, containing a triazolyl-methyl moiety (**5a–m**, entries 3–15), **5j** (entry 12) exhibits the highest affinity with an IC_{50} of 70 nM. This is in the range of *n*-heptyl α-D-mannoside (**1b**), however, compared to the biphenyl derivative **4b**[19] (Fig. 1), this in fact represents a 18-fold reduction of affinity (rIC_{50} 0.06[19] for **4b** vs. rIC_{50} 1.1 for **5j**). At this point, we should recollect that **4b** and **5j** address different docking modes (*out-* and *in-docking-mode*) and therefore also different structural environments.

Antagonists where the linker between the anomeric center and the triazole is extended by an additional carbon (→**6h–k**, entries 16–19) show affinities in the range of 200 nM and therefore - with the exception of 4-pyridyl derivative **6j** (entry 18) - two- to fourfold higher affinity compared to their counterparts with the shorter linker. When the triazole is directly linked to the anomeric center (→**7n–s**, entries 20–25) affinities are 2- to 8-fold reduced, probably as a consequence of the reduced flexibility preventing an optimal

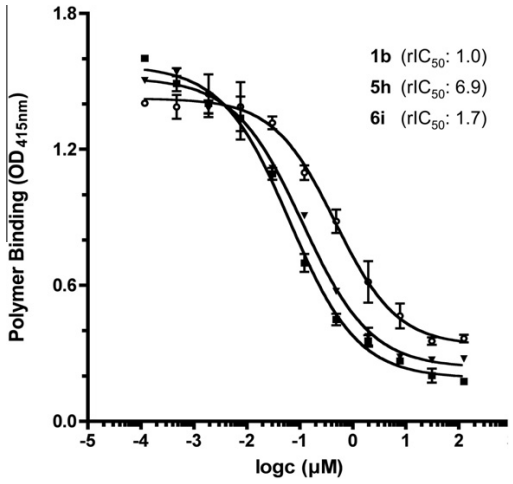


Figure 4. Examples of inhibition curves obtained from the cell-free competitive binding assay.[45] Each assay was run in duplicate and was repeated at least twice. For antagonists **5h**, **6i** and the reference compound **1b** IC_{50} values in the nM range were obtained.

interaction of the aglycone with the tyrosine gate. Finally, the *C*-mannosides **8n**–**s** (entries 26–31), which do not exhibit the exo-anomeric effect of the *O*-mannosides and therefore can more easily adopt the optimal orientation within the tyrosine gate, surprisingly show a twofold reduction in affinity.

2.5. Aggregometry assay

The potential to disaggregate *E. coli* from guinea pig erythrocytes (GPE) was determined for a variety of the mannosylated triazoles in a function-based aggregometry assay.[46] The measurements were performed in triplicates and the corresponding IC_{50} values were calculated by plotting the area under the curve (AUC) of disaggregation against the concentration of the antagonists. *n*-Heptyl α-D-mannopyranoside (**1b**) was again used as reference compound with an IC_{50} of 77.1 μM (Table 2, entry 1). While the antagonists **5e**, **6j**, **6k**, **7o** and **7q** showed IC_{50} values in the range of 200–300 μM, surprisingly no activities could be determined for compounds **5j**, **8q** and **8r** up to a concentration of 700 μM. As earlier observed,[46] the activities obtained from the aggregometry assay are approximately 1000-fold lower than the affinities determined in the target-based competitive assay.

2.6. Conformational analysis of mannosyl triazoles

Compared to their counterparts **7**, where the triazole is directly linked to the anomeric center, most of the *C*-mannosides **8** exhibit a lower affinity. By applying NMR techniques, it was investigated whether this loss of affinity originates from distorted ring conformations. Due to signal overlap, the unprotected mannosides **7** and **8** were not suited for the conformational analysis. However, for the peracetylated derivatives **20n** and **25n** the ring conformation could be assigned based on coupling constants and NOESY experiments. First, the observed ^3J coupling constants for their ring protons were strikingly different. In **20n**, they were in agreement with those expected for a regular $^4C_1$ chair conformation of an α-D-mannopyranosyl ring, with small $J_{1,2}$ and $J_{2,3}$ couplings and large values for $J_{3,4}$ and $J_{4,5}$ (Fig. 5A). In contrast, the large $J_{1,2}$ coupling constant (8.4 Hz) and small to medium values for $J_{2,3}$, $J_{3,4}$ and $J_{4,5}$ found for **25n**, are in agreement with a ring flip of the α-D-mannopyranosyl chair from the common $^4C_1$ to the unusual $^1C_4$ conformation (Fig. 5B). A similar conformational switch has also been observed for α-CF_2-mannosides.[53] As a consequence, the triazolyl-methyl group now is oriented equatorially in *C*-mannoside **25n**, while in **20n** the triazole moiety adopts an axial position.

Subsequent 2D-NOESY measurements (Fig. 5C and D) confirmed this analysis. For both compounds a sequence of seven 2D-NOESY experiments with increasing mixing times from 0.5 s to 2.0 s in steps of 0.25 s was recorded. The intensity of the positive signals grows with increasing mixing time and indicates the relative spatial proximity of a particular proton to that of the source proton. The NOEs of the proton of interest (int_{cross}) were normalized to the intensity of the diagonal peak of the source proton (int_{diag}). Plotting these normalized intensities against the mixing time results in a straight line for each pair of protons. The distances r_{ij} were then calculated from the slopes σ of the linear regression according to $r_{ij} = r_{ref}\,(\sigma_{ref}/\sigma_{ij})^{1/6}$, were r_{ref} is the average distance of the geminal protons H-6a and H-6b, which was chosen as reference (r_{ref} = 1.78 Å).[54,55]

Typically, in the chair conformation of carbohydrates the vicinal proton–proton distances are approx. 2.95 ± 0.15 Å for a diaxial, 2.45 ± 0.15 Å for an axial-equatorial and 2.50 ± 0.20 Å for a diequatorial orientation.[56] As shown in Figure 5A and B, the distances of the ring protons in **20n** and **25n** determined from NOE experiments correlate well with the theoretical values and support the results obtained from the analysis of the coupling constants. In summary, NMR spectroscopic data indicate that the mannopyranosyl chairs in these compounds adopt different conformations, depending on the substituent at C-1.

This conformational analysis offers an explanation for the twofold reduction of affinity found for most of the *C*-mannosides **8** compared to the corresponding *N*-linked triazoles **7**. Due to the inversion of the ring conformation in **8** ($^1C_4$ vs $^4C_1$), the optimal fit into the hydrophilic mannose-binding pocket of FimH is disturbed.

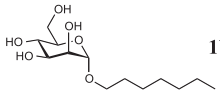
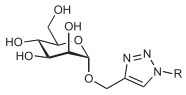
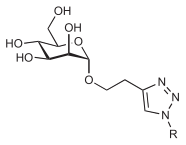
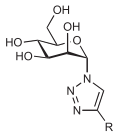
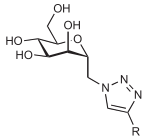
2.7. Pharmacokinetic properties of mannosyl triazoles

Finally, the druglikeness of this new class of FimH antagonists was addressed. For a successful po application in our UTI mouse model,[19] FimH antagonists have to exhibit oral bioavailability, metabolic stability and fast renal elimination to the urinary tract, their place of action. For the evaluation of oral absorption and renal excretion of the triazoles **5–8** physicochemical parameters such as solubility, lipophilicity (distribution coefficients, log $D_{7.4}$) and permeability were determined (Table 1). The mannosides of all four compound families (**5–8**) are all highly soluble (159 μg/mL to > 3 mg/mL) and therefore fulfill a first prerequisite for absorption in the gastrointestinal tract (GIT). All compounds showed low to moderate log $D_{7.4}$ values in the range of < -1.5 to 1.45. While these parameters are beneficial for renal excretion,[57] oral absorption by passive diffusion can only be expected to a minor extent. Indeed, for none of the tested compounds a significant permeation through an artificial membrane (PAMPA,[50] log P_e, P_e: effective permeation) nor membrane retention could be detected. Whereas for a successful oral absorption a log P_e >−5.7 and/or a membrane retention %Mm > 80 % are required,[58] the corresponding values for all triazoles are far from being in this range. Overall, only poor absorption from the GIT can be therefore expected.

3. Conclusions

Crystal structures indicate that the natural ligand oligomannose-3[10] inserts into the tyrosine gate formed by Tyr48, Tyr137 and Ile52 of the carbohydrate recognition domain of FimH (*in-docking-mode*). In contrast, the recently reported high-affinity

Table 2
IC_{50} values of mannosylated triazoles determined in the aggregometry assay[46]

Entry	Ligand		R	Aggregometry assay	
				IC_{50} [μM]	rIC_{50}
1		**1b**[9]		77.1	1.0
2		**5e**		299	3.9
3		**5j**		n.a.	-
5		**6j**		277	3.6
6		**6k**		216	2.8
8		**7o**		289	3.7
9		**7q**		249	3.2
10		**8q**		n.a.	-
11		**8r**		n.a.	-

The relative IC_{50} (rIC_{50}) was calculated by dividing the IC_{50} of the substance of interest by the IC_{50} of the reference compound **1b**. n.a., not active.

biphenyl mannoside **4c** was shown to bind in the *out-docking-mode*, that is, it establishes a π-π-stacking interaction with Tyr48 from the outside of the tyrosine gate.[11] Based on docking studies, we designed a series of low molecular weight mannosyl triazoles, which exhibit an increased conformational flexibility of the aglycone and therefore should allow for binding to the tyrosine gate in the *in-docking-mode*. For their pharmacodynamical evaluation two assay formats, a target-based binding assay[45] and a function-based aggregation assay,[46] were applied. In general, all triazoles **5–8** showed nanomolar affinities, but only one representative, the 4-pyridyl derivative **5j**, was as potent as the reference compound *n*-heptyl mannoside (**1b**). Obviously, the high flexibility of the *n*-heptyl aglycone in **1b** optimally fulfills the spatial requirements of the tyrosine gate. In addition, the hydrophobic contacts established by the substituted triazole aglycone within the tyrosine gate in the *in-docking-mode* are less favorable than the π-π-stacking interaction of biphenyl derivatives[11,19] with Tyr48 in the *out-docking-mode*.

Furthermore, the reduced affinities of the triazolyl-methyl-*C*-mannosides **8** can be rationalized by a disturbed interaction of the mannose moiety. A conformational analysis by ^{1}H NMR and NOESY NMR revealed that in contrast to the other three classes of mannosyl triazoles (compounds **5**, **6** and **7**), the *C*-mannosides **8** do not adopt the common $^4C_1$ but an unusual $^1C_4$ chair conformation, thus preventing an optimal fit of the mannosyl moiety into the hydrophilic mannose-binding pocket of FimH.

Finally, for a successful therapeutic application, FimH antagonists have to exhibit appropriate pharmacokinetic properties, that is, oral bioavailability and fast renal elimination to the urinary tract, their place of action. One prerequisite for absorption in the GIT is sufficient solubility, a property, which is fulfilled by all synthesized antagonists. However, according to their lipophilicity and membrane permeation properties, the mannosyl triazoles are not expected to be orally absorbed. Possible improvements of the pharmacokinetic profiles of mannosyl triazoles are currently studied.

4. Experimental part

4.1. Chemistry

General. NMR spectra were recorded on a Bruker Avance DMX-500 (500 MHz) spectrometer. Assignment of ^{1}H and ^{13}C NMR spectra was achieved using 2D methods (COSY, HSQC). Chemical shifts are expressed in ppm using residual $CHCl_3$ and CD_2HOD as references. Optical rotations were measured using a Perkin-Elmer Polarimeter 341. Electron spray ionization mass

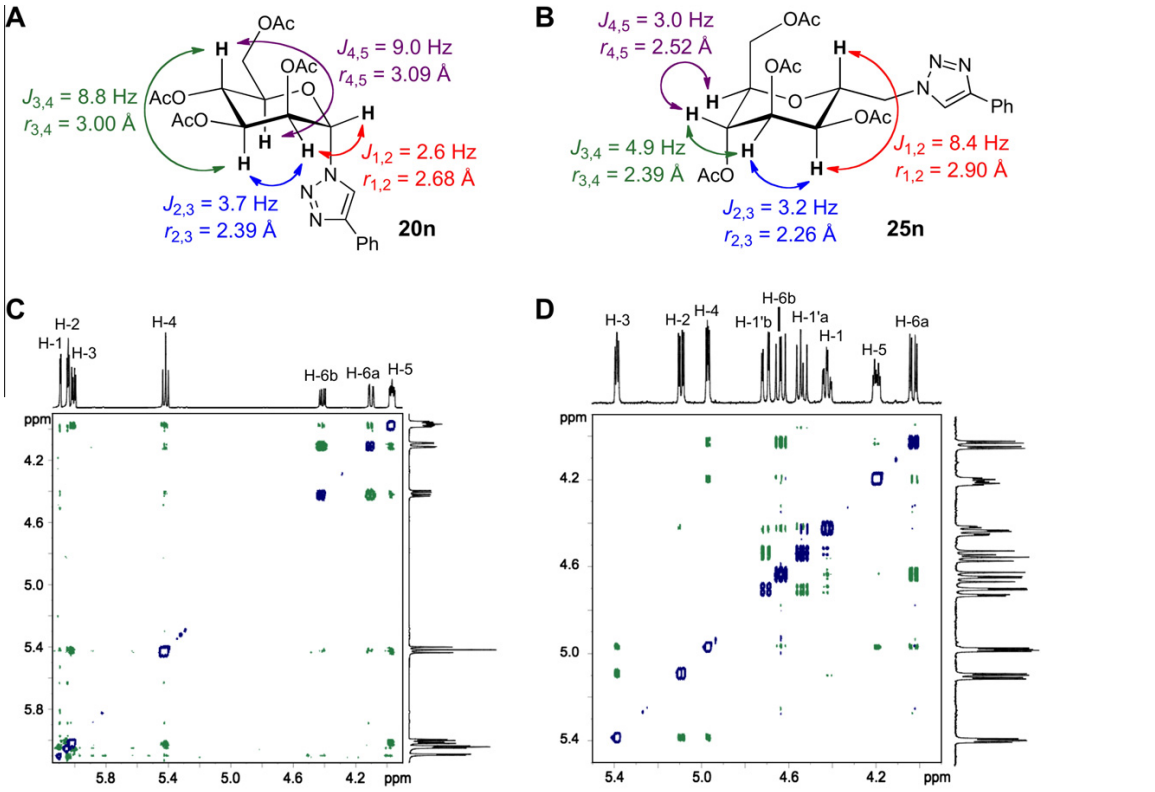


Figure 5. Coupling constants and proton-proton distances for peracetylated triazoles **20n** (A) and **25n** (B) determined by ^{1}H NMR and 2D-NOESY experiments; 2D-NOESY spectra of **20n** (C) and **25n** (D) in $CDCl_3$ with mixing times of 1.5 s (C) and 750 ms (D).

spectra (ESI-MS) were obtained on a Waters micromass ZQ. The HRMS analyses were carried out using a Bruker QTOF. Reactions were monitored by TLC using glass plates coated with silica gel 60 F_{254} (Merck) and visualized by using UV light and/or by charring with a molybdate solution (a 0.02 M solution of ammonium cerium sulfate dihydrate and ammonium molybdate tetrahydrate in aqueous 10% H_2SO_4). MPLC separations were carried out on a Combi-Flash Companion from Teledyne Isco equipped with RediSep normal-phase or C_{18} reversed-phase flash columns. Tetrahydrofurane (THF) was freshly distilled under argon over sodium and benzophenone. Methanol (MeOH) was dried by refluxing with sodium methoxide and distilled immediately before use. Dichloromethane (DCM), ethyl acetate (EtOAc), and toluene were dried by filtration over Al_2O_3 (Fluka, type 5016 A basic).

4.1.1. General procedure A for the synthesis of mannosyl triazoles 5d–i

A mixture of acetylene **14**[37] (1.0 eq), azide **15d–i** (1.5 eq), $CuSO_4{\cdot}5H_2O$ (0.25 eq) and sodium ascorbate (0.5 eq) was dissolved in degassed *tert*-BuOH/H_2O (1:1, 2 mL/0.1 mmol **14**) under argon. After stirring for 1 d the solvents were removed in vacuo and the crude product was first purified by MPLC on silica (DCM/MeOH) and then by reversed-phase chromatography (RP-18, H_2O/MeOH) to yield **5d–i** as colorless solids.

4.1.2. General procedure B for the synthesis of mannosyl triazoles 12a–c, 16j–m and 17h–k

Acetylene **10**[29] or **13**[36] (1.0 eq) and azide **11a–c** or **15h–m** (1.5–2 eq) were dissolved in THF/*tert*-BuOH/H_2O (1:1:1, 1.5 mL/0.1 mmol **10** or **11**). The mixture was degassed in an ultrasound bath under a flow of argon for 20 min. Then 0.5 M aq $CuSO_4{\cdot}5H_2O$ (0.25 eq) and 1 M aq sodium ascorbate (0.5 eq) were added under argon at rt. After stirring overnight the solvents were removed in vacuo and the crude product was purified by MPLC on silica (petrol ether/EtOAc) to yield **12a–c**, **16j–m** and **17h–k** as colorless oils.

4.1.3. General procedure C for the synthesis of mannosyl triazoles 20n–s and 25n–s

Azide **18**[42] or **24** (1.0 eq) and acetylene **19n–s** (2.0 eq) were dissolved in THF/*tert*-BuOH/H_2O (1:1:1, 3 mL/0.1 mmol **18** or **24**). The mixture was degassed in an ultrasound bath under a flow of argon for 20 min. Then 0.2 M aq $CuSO_4{\cdot}5H_2O$ (0.2 eq) and 1 M aq sodium ascorbate (0.4 eq) were added under argon at rt. After stirring for 1–2 d the solvents were removed in vacuo and the crude product was purified by MPLC on silica (petrol ether/EtOAc) to yield **20n–s** and **25n–s** as colorless oils.

4.1.4. General procedure D for deacetylation

To a solution of the acetylated compound (38–50 mg) in MeOH (3 mL) was added 1 M NaOMe/MeOH (0.3 mL). The mixture was stirred at rt for 3–6 h. The solution was concentrated and the residue was purified by MPLC on reversed phase (RP-18 column, H_2O/MeOH) and P2 size-exclusion chromatography to afford the target molecule as a colorless solid after a final lyophilization from water/dioxane.

4.1.5. Synthesis of azide 24

4.1.5.1. (2,3,4,6-Tetra-*O*-acetyl-α-D-mannopyranosyl)-*N*-*tert*-butoxycarbonyl-methylamine (22). Cyanide **21**[43] (1.63 g, 4.57 mmol), Boc_2O (1.49 g, 6.86 mmol) and Pd/C (10%, 250 mg) were suspended in EtOAc (25 mL) and hydrogenated (4 bar H_2) at rt for 4 h. After filtration over Celite, fresh Pd/C (10%, 750 mg)

was added and the mixture was hydrogenated (4 bar H_2) for additional 17 h. The suspension was filtered over Celite and concentrated. The residue was purified by MPLC on silica (petrol ether/EtOAc) to give **22** (1.51 g, 72%) as a colorless solid.

^{1}H NMR (500 MHz, $CDCl_3$): δ 1.44 (s, 9H, $C(CH_3)_3$), 2.07, 2.10, 2.10 (3 s, 12H, 4 $COCH_3$), 3.38 (m, 2H, H-1′), 3.99–4.05 (m, 2H, H-1, H-5), 4.07 (dd, J = 3.8, 11.8 Hz, 1H, H-6a), 4.54 (dd, J = 6.9, 11.7 Hz, 1H, H-6b), 4.79 (m, 1H, NH), 5.07 (dd, J = 5.3, 6.4 Hz, 1H, H-4), 5.10 (dd, J = 3.3, 6.0 Hz, 1H, H-2), 5.26 (dd, J = 3.3, 6.5 Hz, 1H, H-3); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.70, 20.73, 20.76, 20.81 (4 $COCH_3$), 28.3 ($C(CH_3)_3$), 39.7 (C-1′), 61.2 (C-6), 67.50, 67.51 (C-2, C-4), 68.0 (C-3), 71.1 (C-1), 72.2 (C-5), 79.7 ($C(CH_3)_3$), 155.8 (NCO), 169.5, 169.6, 169.9, 170.7 (4 $COCH_3$); ESI-MS Calcd for $C_{20}H_{31}NNaO_{11}$ $[M+Na]^+$: 484.18, Found: 484.11.

4.1.5.2. *N-tert*-Butoxycarbonyl-(α-D-mannopyranosyl)methylamine (23). A solution of **22** (1.47 g, 3.19 mmol) in MeOH (20 mL) was treated with 1 M methanolic NaOMe (2 mL) under argon at rt for 3 h. The reaction mixture was neutralized with acetic acid and concentrated. The residue was purified by MPLC on silica (DCM/MeOH) to give **23** (925 mg, 99%) as a colorless solid.

^{1}H NMR (500 MHz, CD_3OD): δ 1.44 (s, 9H, $C(CH_3)_3$), 3.32 (m, 2H, H-1′), 3.54 (m, 1H, H-5), 3.63 (t, J = 7.6 Hz, 1H, H-4), 3.68 (dd, J = 3.1, 7.8 Hz, 1H, H-3), 3.74 (dd, J = 2.8, 11.8 Hz, 1H, H-6a), 3.77 (m, 1H, H-2), 3.79 (dd, J = 6.4, 11.8 Hz, 1H, H-6b), 3.86 (m, 1H, H-1), 6.72 (m, 1H, NH); ^{13}C NMR (125 MHz, CD_3OD): δ 28.8 ($C(CH_3)_3$), 40.6 (C-1′), 62.7 (C-6), 69.6 (C-4), 70.2 (C-2), 72.7 (C-3), 76.9 (C-1), 77.4 (C-5), 80.2 ($C(CH_3)_3$), 158.6 (NCO); ESI-MS Calcd for $C_{12}H_{23}NNaO_7$ $[M+Na]^+$: 316.14, Found: 316.03.

4.1.5.3. (2,3,4,6-Tetra-*O*-acetyl-α-D-mannopyranosyl)methylazide (24). *Triflyl azide stock solution preparation*:[44] Sodium azide (796 mg, 12.2 mmol) was dissolved in water (2 mL). Toluene (2 mL) was added, and the mixture was cooled to 0 °C with stirring. Then triflic anhydride (1.31 mL, 6.12 mmol) was added dropwise. The biphasic reaction mixture was stirred vigorously at 0 °C for 30 min and at 10 °C for another 2 h. The reaction mixture was neutralized with satd aq $NaHCO_3$. The phases were separated, and the aqueous phase extracted with toluene (2 × 2 mL). The organic layers were combined to give the triflyl azide stock solution. *Amine-azide exchange*: A solution of **23** (430 mg, 1.47 mmol) in dioxane/water (2:1, 15 mL) was treated with concentrated HCl (5 mL) under argon at rt for 4 h. The mixture was concentrated and the residue was dried in high vacuo. Then, the crude amine hydrochloride (341 mg), $NaHCO_3$ (492 mg, 5.86 mmol) and $CuSO_4{\cdot}5H_2O$ (14.1 mg, 61 μmol) were dissolved in water (1.91 mL). The triflyl azide stock solution (3.25 mL, 3.3 mmol) was added and the biphasic reaction mixture was made homogenous by the addition of MeOH (12.6 mL). The mixture was stirred at rt for 20 h. The solvents were removed in vacuo and the residue was taken up in dry pyridine (10 mL), and acetic anhydride (4 mL) was added. The reaction mixture was stirred at rt under argon for 4 h. The solvents were removed in vacuo and the crude product was purified by MPLC on silica (petrol ether/EtOAc) to yield **24** (459 mg, 81%) as a colorless oil.

IR (film) 2102 (vs, N_3), 1747 (vs, CO) cm^{-1}; ^{1}H NMR (500 MHz, $CDCl_3$): δ 2.04, 2.07, 2.08, 2.10 (4 s, 12H, 4 $COCH_3$), 3.29 (dd, J = 3.3, 13.4 Hz, 1H, H-1′a), 3.49 (dd, J = 7.3, 13.4 Hz, 1H, H-1′b), 4.05–4.09 (m, 2H, H-5, H-6a), 4.15 (dt, J = 3.2, 7.1 Hz, 1H, H-1), 4.60 (m, 1H, H-6b), 5.01 (dd, J = 4.4, 6.0 Hz, 1H, H-4), 5.14 (dd, J = 3.4, 6.9 Hz, 1H, H-2), 5.27 (dd, J = 3.4, 6.0 Hz, 1H, H-3); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.61, 20.63, 20.65, 20.74 (4 $COCH_3$), 50.1 (C-1′), 60.8 (C-6), 67.0 (C-2), 67.5 (C-3), 67.6 (C-4), 70.5 (C-1), 72.9 (C-5), 169.3, 169.5, 169.6, 170.6 (4 $COCH_3$); ESI-MS Calcd for $C_{15}H_{21}N_3NaO_9$ $[M+Na]^+$: 410.12, Found: 410.04.

4.1.6. Synthesis of peracetylated mannosyl triazoles

4.1.6.1. Methyl 4-[4-((2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyloxy)methyl)-1*H*-1,2,3-triazol-1-yl]-benzoate (12a). Following general procedure B, **10** (40.0 mg, 0.103 mmol) was reacted with methyl 4-azidobenzoate (**11a**,[30] 36.5 mg, 0.206 mmol), 0.5 M $CuSO_4$ (52 μL, 26 μmol) and 1 M sodium ascorbate (52 μL, 52 μmol) to yield **12a** (55.8 mg, 96%).

$[\alpha]_D$ +45.0 (*c* 1.03, $CHCl_3$); IR (film) 1747 (vs, CO) cm^{-1}; ^{1}H NMR (500 MHz, $CDCl_3$): δ 1.99, 2.04, 2.12, 2.16 (4 s, 12H, 4 $COCH_3$), 3.97 (s, 3H, OMe), 4.10 (ddd, J = 2.2, 5.1, 9.5 Hz, 1H, H-5), 4.14 (dd, J = 2.3, 12.2 Hz, 1H, H-6a), 4.32 (dd, J = 5.2, 12.2 Hz, 1H, H-6b), 4.79, 4.95 (A, B of AB, J = 12.6 Hz, 2H, H-1′), 5.01 (d, J = 1.1 Hz, 1H, H-1), 5.28 (dd, J = 1.7, 3.1 Hz, 1H, H-2), 5.32 (t, J = 9.8 Hz, 1H, H-4), 5.36 (dd, J = 3.2, 10.0 Hz, 1H, H-3), 7.88 (AA′ of AA′BB′, J = 8.7 Hz, 2H, C_6H_4), 8.11 (s, 1H, C_2N_3H), 8.23 (BB′ of AA′BB′, J = 8.7 Hz, 2H, C_6H_4); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.7, 20.8, 20.9 (4C, 4 $COCH_3$), 52.5 (OMe), 61.0 (C-1′), 62.4 (C-6), 66.0 (C-4), 68.8, 69.0, 69.4 (C-2, C-3, C-5), 97.0 (C-1), 120.0 (2C, C_6H_4), 121.0 (C_2N_3H–C5), 130.4 (C_6H_4–C1), 131.4 (2C, C_6H_4), 139.9 (C_6H_4–C4), 144.7 (C_2N_3H–C4), 169.7, 170.0, 170.1, 170.7 (5C, 5 CO); ESI-MS Calcd for $C_{25}H_{29}N_3NaO_{12}$ $[M+Na]^+$: 586.02, Found: 586.16.

4.1.6.2. Ethyl 3-[4-((2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyloxy)methyl)-1*H*-1,2,3-triazol-1-yl]-benzoate (12b). Following general procedure B, **10** (50.0 mg, 0.129 mmol) was reacted with ethyl 3-azidobenzoate (**11b**,[31] 49.3 mg, 0.258 mmol), 0.5 M $CuSO_4$ (64 μL, 32 μmol) and 1 M sodium ascorbate (64 μL, 64 μmol) to yield **12b** (72.7 mg, 97%).

$[\alpha]_D$ +40.2 (*c* 1.04, $CHCl_3$); IR (film) 1749 (vs, CO) cm^{-1}; ^{1}H NMR (500 MHz, $CDCl_3$): δ 1.43 (t, J = 7.2 Hz, 1H, CH_3), 1.99, 2.04, 2.12, 2.16 (4 s, 12H, 4 $COCH_3$), 4.09–4.15 (m, 2H, H-5, H-6a), 4.32 (dd, J = 5.0, 12.1 Hz, 1H, H-6b), 4.44 (q, J = 7.2 Hz, 2H, OCH_2), 4.79, 4.95 (A, B of AB, J = 12.4 Hz, 2H, H-1′), 5.02 (d, J = 1.4 Hz, 1H, H-1), 5.28 (dd, J = 1.7, 3.2 Hz, 1H, H-2), 5.31 (m, 1H, H-4), 5.36 (dd, J = 3.3, 9.9 Hz, 1H, H-3), 7.64 (t, J = 8.0 Hz, 1H, C_6H_4–H5), 8.02 (ddd, J = 0.9, 2.1, 8.0 Hz, 1H, C_6H_4–H4), 8.11 (s, 1H, C_2N_3H), 8.14 (d, J = 7.9 Hz, 1H, C_6H_4-H6), 8.38 (t, J = 1.7 Hz, 1H, C_6H_4-H2); ^{13}C NMR (125 MHz, $CDCl_3$): δ 14.3 (CH_3), 20.67, 20.69, 20.78, 20.88 (4 $COCH_3$), 60.9 (C-1′), 61.7 (OCH_2), 62.4 (C-6), 66.0 (C-4), 68.8 (C-3), 69.0 (C-5), 69.4 (C-2), 96.9 (C-1), 121.2 (C_6H_4), 121.3 (C_2N_3H–C5), 124.8, 129.9, 130.0, 132.3, 137.0 (C_6H_4), 144.5 (C_2N_3H–C4), 165.2, 169.7, 169.9, 170.1, 170.7 (5 CO); ESI-MS Calcd for $C_{26}H_{32}N_3O_{12}$ $[M+H]^+$: 578.20, Found: 578.19.

4.1.6.3. Methyl 5-[4-((2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyloxy)methyl)-1*H*-1,2,3-triazol-1-yl]-nicotinate (12c). Following general procedure B, **10** (40.0 mg, 0.103 mmol) was reacted with methyl 5-azidonicotinate (**11c**,[32] 32.5 mg, 0.182 mmol), 0.5 M $CuSO_4$ (52 μL, 26 μmol) and 1 M sodium ascorbate (52 μL, 52 μmol). The crude product was dissolved in DCM (10 mL) and washed with 0.1 M aq EDTA (5 mL). The aqueous phase was extracted with DCM (2 × 10 mL) and the combined organic layers were dried with Na_2SO_4 and evaporated to dryness. The residue was purified by MPLC on silica (petrol ether/EtOAc) to give **12c** (42.4 mg, 73%).

$[\alpha]_D$ +39.7 (*c* 1.06, $CHCl_3$); IR (film) 1733 (vs, CO) cm^{-1}; ^{1}H NMR (500 MHz, $CDCl_3$): δ 1.98, 2.03, 2.11, 2.15 (4 s, 12H, 4 $COCH_3$), 4.01 (s, 3H, OCH_3), 4.09 (m, 1H, H-5), 4.14 (dd, J = 2.4, 12.2 Hz, 1H, H-6a), 4.31 (dd, J = 5.2, 12.3 Hz, 1H, H-6b), 4.79, 4.96 (A, B of AB, J = 12.5 Hz, 2H, H-1′), 5.01 (d, J = 1.4 Hz, 1H, H-1), 5.27 (dd, J = 1.7, 3.0 Hz, 1H, H-2), 5.30 (t, J = 9.8 Hz, 1H, H-4), 5.34 (dd, J = 3.3, 9.9 Hz, 1H, H-3), 8.17 (s, 1H, C_2N_3H), 8.69 (t, J = 2.0 Hz, 1H, C_5H_3N-H2), 9.27, 9.30 (2 s, 2H, C_5H_3N-H4, H6); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.63, 20.65, 20.75, 20.84 (4 $COCH_3$), 52.9 (OMe), 60.8 (C-1′), 62.4 (C-6), 66.0 (C-4), 68.8, 68.9 (C-3, C-5), 69.3 (C-2), 97.0 (C-1), 121.2 (C_2N_3H–C5), 126.9, 128.6, 133.3

(C_5H_3N), 145.1 (2C, C_5H_3N, (C_2N_3H–C4), 150.7 (C_5H_3N), 164.4, 169.7, 169.9, 170.1, 170.7 (5 CO); ESI-MS Calcd for $C_{24}H_{29}N_4O_{12}$ $[M+H]^+$: 565.18, Found: 565.15.

4.1.6.4. [1-(Pyridin-4-yl)-1,2,3-triazol-4-yl]methyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (16j). Following general procedure B, **10** (100 mg, 0.26 mmol) was reacted with 4-azidopyridine (**15j**,[41] 47 mg, 0.39 mmol), 0.5 M $CuSO_4$ (130 μL, 65 μmol) and 1 M sodium ascorbate (130 μL, 130 μmol). The crude product was dissolved in DCM (20 mL) and washed with 0.1 M aq EDTA (10 mL). The aqueous phase was extracted with DCM (2 × 20 mL) and the combined organic layers were dried with Na_2SO_4 and evaporated to dryness. The residue was purified by MPLC on silica (DCM/MeOH) to give **16j** (114 mg, 87%).

$[\alpha]_D$ +52.1 (*c* 2.26, DCM); IR (film) 1748 (vs, CO) cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$): δ 1.92, 1.97, 2.05, 2.09 (4s, 12H, 4 $COCH_3$), 3.96–4.14 (m, 2H, H-5, H-6a), 4.24 (dd, *J* = 4.8, 12.1 Hz, 1H, H-6b), 4.73, 4.89 (A, B of AB, *J* = 12.5 Hz, 2H, H-1′), 4.94 (s, 1H, H-1), 5.14–5.33 (m, 3H, H-2, H-3, H-4), 7.70 (m, 2H, C_5H_4N), 8.15 (s, 1H, C_2N_3H), 8.73 (m, 2H, C_5H_4N); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.77, 20.78, 20.88, 20.96 (4 $COCH_3$), 61.0 (C-1′), 62.5 (C-6), 66.1, 68.9, 69.0, 69.5 (C-2, C-3, C-4, C-5), 97.1 (C-1), 113.9 (2C, C_5H_4N), 120.7 (C_2N_3H–C5), 143.0 (C_5H_4N–C1), 145.2 (C_2N_3H–C4), 151.9 (2C, C_5H_4N), 169.8, 170.1, 170.2, 170.8 (4 $COCH_3$); ESI-MS Calcd for $C_{22}H_{26}N_4NaO_{10}$ $[M+Na]^+$: 529.16, Found: 529.07.

4.1.6.5. [1-(4′-Fluorophenyl)-1,2,3-triazol-4-yl]methyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (16k). Following general procedure B, **10** (100 mg, 0.26 mmol) was reacted with 1-azido-4-fluorobenzene (**15k**,[41] 53 mg, 0.39 mmol), 0.5 M $CuSO_4$ (130 μL, 65 μmol) and 1 M sodium ascorbate (130 μL, 130 μmol) to yield **16k** (127 mg, 94%).

$[\alpha]_D$ +42.0 (*c* 1.00, DCM); IR (film) 1749 (vs, CO) cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$): δ 1.96, 2.01, 2.09, 2.13 (4s, 12H, 4 $COCH_3$), 4.07 (m, 1H, H-5), 4.11 (dd, *J* = 2.4, 12.2 Hz, 1H, H-6a), 4.28 (dd, *J* = 5.2, 12.2 Hz, 1H, H-6b), 4.75, 4.91 (A, B of AB, *J* = 12.4 Hz, 2H, H-1′), 4.98 (d, *J* = 1.6 Hz, 1H, H-1), 5.24–5.31 (m, 2H, H-2, H-4), 5.33 (dd, *J* = 3.3, 10.0 Hz, 1H, H-3), 7.21 (m, 2H, C_6H_4), 7.71 (m, 2H, C_6H_4), 7.97 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.87, 20.88, 20.98, 21.07 (4 $COCH_3$), 61.3 (C-1′), 62.6 (C-6), 66.3, 69.0, 69.2, 69.7 (C-2, C-3, C-4, C-5), 97.2 (C-1), 117.0 (d, *J* = 22.5 Hz, 2C, C_6H_4), 121.6 (C_2N_3H–C5), 122.9 (d, *J* = 8.8 Hz, 2C, C_6H_4), 133.4 (d, *J* = 3.8 Hz, C_6H_4–C1), 145.0 (C_2N_3H–C4), 163.8 (d, *J* = 247.5 Hz, C_6H_4–C4), 169.9, 170.2, 170.3, 170.9 (4 $COCH_3$); ESI-MS Calcd for $C_{23}H_{26}FN_3NaO_{10}$ $[M+Na]^+$: 546.16, Found: 546.15.

4.1.6.6. [1-(3′-Fluorophenyl)-1,2,3-triazol-4-yl]methyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (16l). Following general procedure B, **10** (100 mg, 0.26 mmol) was reacted with 1-azido-3-fluorobenzene (**15l**,[41] 53 mg, 0.39 mmol), 0.5 M $CuSO_4$ (130 μL, 65 μmol) and 1 M sodium ascorbate (130 μL, 130 μmol) to yield **16l** (115 mg, 85%).

$[\alpha]_D$ +47.5 (*c* 2.14, DCM); IR (film) 1748 (vs, CO) cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$): δ 1.96, 2.01, 2.09, 2.13 (4s, 12H, 4 $COCH_3$), 4.02–4.15 (m, 2H, H-5, H-6a), 4.29 (dd, *J* = 5.1, 12.2 Hz, 1H, H-6b), 4.75, 4.91 (A, B of AB, *J* = 12.5 Hz, 2H, H-1′), 4.98 (s, 1H, H-1), 5.22–5.36 (m, 3H, H-2, H-3, H-4), 7.14 (m, 1H, C_6H_4), 7.44–7.59 (m, 3H, C_6H_4), 8.02 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.86, 20.88, 20.97, 21.07 (4 $COCH_3$), 61.2 (C-1′), 62.6 (C-6), 66.2, 69.0, 69.2, 69.6 (C-2, C-3, C-4, C-5), 97.2 (C-1), 108.6 (d, *J* = 26.3 Hz, C_6H_4), 116.0 (d, *J* = 17.5 Hz, C_6H_4), 116.1 (m, C_6H_4), 121.3 (C_2N_3H–C5), 131.5 (d, *J* = 8.8 Hz, C_6H_4), 138.2 (d, *J* = 10.0 Hz, C_6H_4–C1), 144.7 (C_2N_3H–C4), 163.3 (d, *J* = 247.5 Hz, C_6H_4–C4), 169.9, 170.1, 170.3, 170.9 (4 $COCH_3$); ESI-MS Calcd for $C_{23}H_{26}FN_3NaO_{10}$ $[M+Na]^+$: 546.16, Found: 546.15.

4.1.6.7. [1-(4′-Methoxyphenyl)-1,2,3-triazol-4-yl]methyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (16m). Following general procedure B, **10** (100 mg, 0.26 mmol) was reacted with 1-azido-4-methoxybenzene (**15m**,[40] 58 mg, 0.39 mmol), 0.5 M $CuSO_4$ (130 μL, 65 μmol) and 1 M sodium ascorbate (130 μL, 130 μmol) to yield **16m** (128 mg, 92%).

$[\alpha]_D$ +45.0 (*c* 2.26, DCM); IR (film) 1749 (vs, CO) cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$): δ 1.93, 1.98, 2.06, 2.10 (4s, 12H, 4 $COCH_3$), 3.81 (s, 3H, OCH_3), 3.99–4.12 (m, 2H, H-5, H-6a), 4.26 (dd, *J* = 5.3, 12.4 Hz, 1H, H-6b), 4.71, 4.87 (A, B of AB, *J* = 12.4 Hz, 2H, H-1′), 4.95 (d, *J* = 1.4 Hz, 1H, H-1), 5.20–5.28 (m, 2H, H-2, H-4), 5.30 (dd, *J* = 3.2, 10.0 Hz, 1H, H-3), 6.97 (m, 2H, C_6H_4), 7.58 (m, 2H, C_6H_4), 7.91 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.78, 20.80, 20.89, 20.98 (4 $COCH_3$), 55.7 (OCH_3), 61.2 (C-1′), 62.5 (C-6), 66.1, 68.9, 69.1, 69.5 (C-2, C-3, C-4, C-5), 97.0 (C-1), 114.9 (2C, C_6H_4), 121.5 (C_2N_3H–C5), 122.4 (2C, C_6H_4), 130.4 (C_6H_4–C1), 144.1 (C_2N_3H–C4), 160.0 (C_6H_4–C4), 169.8, 170.0, 170.2, 170.8 (4 $COCH_3$); ESI-MS Calcd for $C_{24}H_{29}N_3NaO_{11}$ $[M+Na]^+$: 558.18, Found: 558.18.

4.1.6.8. [1-(3′-Methoxybenzyl)-1,2,3-triazol-4-yl]ethyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (17h). Following general procedure B, **13** (55 mg, 0.14 mmol) was reacted with 3-methoxybenzylazide (**15h**,[39] 34 mg, 0.21 mmol), 0.5 M $CuSO_4$ (70 μL, 35 μmol) and 1 M sodium ascorbate (70 μL, 70 μmol) to yield **17h** (65 mg, 83%).

$[\alpha]_D$ +36.5 (*c* 1.00, DCM); IR (film) 1748 (vs, CO) cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$): δ 1.85, 1.88, 1.93, 1.99 (4s, 12H, 4 $COCH_3$), 2.86 (tt, *J* = 4.8, 9.9 Hz, 2H, H-2′), 3.56 (dt, *J* = 6.6, 9.5 Hz, 1H, H-1′a), 3.63 (s, 3H, OCH_3), 3.69 (m, 1H, H-5), 3.83 (dt, *J* = 6.6, 9.5 Hz, 1H, H-1′b), 3.91 (dd, *J* = 2.4, 12.3 Hz, 1H, H-6a), 4.09 (dd, *J* = 5.2, 12.3 Hz, 1H, H-6b), 4.65 (d, *J* = 1.7 Hz, 1H, H-1), 5.03 (dd, *J* = 1.8, 3.0 Hz, 1H, H-2), 5.07–5.16 (m, 2H, H-3, H-4), 5.33 (s, 2H, CH_2Ar), 6.68, 6.72, 7.13 (m, 4H, C_6H_4), 7.68 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.85, 20.88, 21.02 (4C, 4 $COCH_3$), 26.4 (C-2′), 54.1 (CH_2Ar), 55.4 (OCH_3), 62.5 (C-6), 66.2 (C-4), 67.2 (C-1′), 68.9, 69.2, 69.7 (C-2, C-3, C-5), 97.6 (C-1), 113.9, 114.2, 120.4, 130.3 (C_6H_4, C_2N_3H–C5), 136.5 (C_6H_4–C1), 144.8 (C_2N_3H–C4), 160.2 (C_6H_4–C3), 169.8, 170.1, 170.2, 170.8 (4 $COCH_3$); ESI-MS Calcd for $C_{26}H_{33}N_3NaO_{11}$ $[M+Na]^+$: 586.21, Found 586.29.

4.1.6.9. [1-(4′-Nitrophenyl)-1,2,3-triazol-4-yl]ethyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (17i). Following general procedure B, **13** (55 mg, 0.14 mmol) was reacted with 1-azido-4-nitrobenzene (**15i**,[40] 34 mg, 0.21 mmol), 0.5 M $CuSO_4$ (70 μL, 35 μmol) and 1 M sodium ascorbate (70 μL, 70 μmol) to yield **17i** (74 mg, 94%).

$[\alpha]_D$ +28.1 (*c* 1.00, DCM); IR (film) 1748 (vs, CO) cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$): δ 1.92, 1.99, 2.06, 2.11 (4s, 12H, 4 $COCH_3$), 3.12 (m, 2H, H-2′), 3.69–3.82 (m, 2H, H-5, H-1′a), 3.99–4.09 (m, 2H, H-6a, H-1′b), 4.19 (dd, *J* = 5.3, 12.3 Hz, 1H, H-6b), 4.84 (d, *J* = 1.6 Hz, 1H, H-1), 5.17–5.27 (m, 2H, H-2, H-4), 5.30 (dd, *J* = 3.3, 10.2 Hz, 1H, H-3), 8.03 (s, 1H, C_2N_3H), 8.06, 8.37 (m, 4H, C_6H_4); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.75, 20.89, 20.91, 21.02 (4 $COCH_3$), 26.3 (C-2′), 62.5 (C-6), 66.1 (C-4), 66.6 (C-1′), 69.1, 69.2, 69.6 (C-2, C-3, C-5), 97.5 (C-1), 120.6 (C_2N_3H–C5), 120.6 (2C, C_6H_4), 125.6 (2C, C_6H_4), 141.5 (C_6H_4–C1), 146.2 (C_2N_3H–C4), 147.2 (C_6H_4–C4), 169.7, 170.3, 170.4, 170.8 (4 $COCH_3$); ESI-MS Calcd for $C_{26}H_{33}N_3NaO_{11}$ $[M+Na]^+$: 587.17, Found 587.25.

4.1.6.10. [1-(Pyridin-4′-yl)-1,2,3-triazol-4-yl]ethyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (17j). Following general procedure B, **13** (60 mg, 0.15 mmol) was reacted with 4-azidopyridine (**15j**,[41] 28 mg, 0.23 mmol), 0.5 M $CuSO_4$ (75 μL, 38 μmol) and 1 M sodium ascorbate (75 μL, 75 μmol). The crude product was dissolved in DCM (20 mL) and washed with 0.1 M aq EDTA

(10 mL). The aqueous phase was extracted with DCM (2 × 20 mL) and the combined organic layers were dried with Na_2SO_4 and evaporated to dryness. The residue was purified by MPLC on silica (petrol ether/EtOAc) to give **17j** (74 mg, 94%).

$[\alpha]_D$ +32.6 (*c* 0.99, DCM); IR (film) 1748 (vs, CO) cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$): δ 1.88, 1.97, 2.05, 2.10 (4s, 12H, 4 $COCH_3$), 3.10 (m, 2H, H-2′), 3.64–3.79 (m, 2H, H-5, H-1′a), 3.96–4.06 (m, 2H, H-6a, H-1′b), 4.17 (dd, *J* = 5.3, 12.3 Hz, 1H, H-6b), 4.83 (d, *J* = 1.6 Hz, 1H, H-1), 5.16–5.26 (m, 2H, H-2, H-4), 5.28 (dd, *J* = 3.4, 10.1 Hz, 1H, H-3), 7.78 (dd, *J* = 1.6, 4.7 Hz, 2H, C_5H_4N), 8.04 (s, 1H, C_2N_3H), 8.72 (dd, *J* = 1.4, 4.9 Hz, 2H, C_5H_4N); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.68, 20.88, 21.0 (4C, 4 $COCH_3$), 26.3 (C-2′), 62.5 (C-6), 66.0 (C-4), 66.5 (C-1′), 69.0, 69.2, 69.7 (C-2, C-3, C-5), 97.4 (C-1), 113.9 (2C, C_5H_4N), 120.0 (C_2N_3H–C5), 143.3 (C_5H_4N–C1), 146.1 (C_2N_3H–C4), 151.8 (2C, C_5H_4N), 169.7, 170.3, 170.4, 170.8 (4 $COCH_3$); ESI-MS Calcd for $C_{26}H_{33}N_3NaO_{11}$ $[M+Na]^+$: 543.18, Found: 543.14.

4.1.6.11. [1-(4′-Fluorophenyl)-1,2,3-triazol-4-yl]ethyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (17k). Following general procedure B, **13** (60 mg, 0.15 mmol) was reacted with 1-azido-4-fluorobenzene (**15k**,[41] 32 mg, 0.23 mmol), 0.5 M $CuSO_4$ (77 μL, 38 μmol) and 1 M sodium ascorbate (75 μL, 75 μmol) to yield **17k** (77 mg, 96%).

$[\alpha]_D$ +32.0 (*c* 1.01, DCM); IR (film) 1751 (vs, CO) cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$): δ 1.90, 1.96, 2.05, 2.10 (4s, 12H, 4 $COCH_3$), 3.07 (m, 2H, H-2′), 3.67–3.79 (m, 2H, H-5, H-1′a), 3.96–4.06 (m, 2H, H-6a, H-1′b), 4.18 (dd, *J* = 5.2, 12.3 Hz, 1H, H-6b), 4.82 (d, *J* = 1.6 Hz, 1H, H-1), 5.16–5.24 (m, 2H, H-2, H-4), 5.28 (dd, *J* = 3.4, 10.1 Hz, 1H, H-3), 7.16, 7.73 (m, 4H, C_6H_4), 7.85 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.71, 20.85, 20.88, 21.01 (4 $COCH_3$), 26.3 (C-2′), 62.5 (C-6), 66.1 (C-4), 66.8 (C-1′), 68.9, 69.2, 69.7 (C-2, C-3, C-5), 97.4 (C-1), 116.8 (d, *J* = 22.5 Hz, 2C, C_6H_4), 120.7 (C_2N_3H–C5), 122.6 (d, *J* = 8.8 Hz, 2C, C_6H_4), 133.6 (d, *J* = 2.5 Hz, C_6H_4–C1), 145.4 (C_2N_3H–C4), 162.5 (d, *J* = 247.5 Hz, C_6H_4–C4), 169.7, 170.2, 170.3, 170.8 (4 $COCH_3$); ESI-MS Calcd for $C_{26}H_{33}N_3NaO_{11}$ $[M+Na]^+$: 560.18, Found 560.17.

4.1.6.12. 1-(2,3,4,6-Tetra-*O*-acetyl-α-D-mannopyranosyl)-4-phenyl-1,2,3-triazole (20n). Following general procedure C, **18** (102 mg, 0.273 mmol) was reacted with phenylacetylene (**19n**, 60 μL, 0.55 mmol), 0.2 M $CuSO_4$ (273 μL, 54.6 μmol) and 1 M sodium ascorbate (109 μL, 109 μmol) to yield **20n** (110 mg, 84%).

$[\alpha]_D$ +65.5 (*c* 1.01, $CHCl_3$); ^1H NMR (500 MHz, $CDCl_3$): δ 2.07, 2.07, 2.09, 2.20 (4 s, 12H, 4 $COCH_3$), 3.94 (ddd, *J* = 2.5, 5.4, 9.0 Hz, 1H, H-5), 4.08 (dd, *J* = 2.5, 12.5 Hz, 1H, H-6a), 4.39 (dd, *J* = 5.4, 12.5 Hz, 1H, H-6b), 5.39 (t, *J* = 8.9 Hz, 1H, H-4), 5.98 (dd, *J* = 3.7, 8.8 Hz, 1H, H-3), 6.02 (dd, *J* = 2.7, 3.7 Hz, 1H, H-2), 6.07 (d, *J* = 2.6 Hz, 1H, H-1), 7.32–7.39, 7.44–7.47, 7.85–7.87 (m, 5H, C_6H_5), 7.96 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.5, 20.6, 20.6, 20.7 (4 $COCH_3$), 61.5 (C-6), 66.0 (C-4), 68.2 (C-2), 68.7 (C-3), 72.1 (C-5), 83.5 (C-1), 119.7 (C_2N_3H–C5), 125.8, 128.6, 128.9, 129.6 (6C, C_6H_5), 148.2 (C_2N_3H–C4), 169.2, 169.6, 169.6, 170.4 (4 $COCH_3$); ESI-MS Calcd for $C_{22}H_{25}N_3NaO_9$ $[M+Na]^+$: 498.15, Found: 498.20.

4.1.6.13. 1-(2,3,4,6-Tetra-*O*-acetyl-α-D-mannopyranosyl)-4-(4-methylphenyl)-1,2,3-triazole (20o). Following general procedure C, **18** (50 mg, 0.13 mmol) was reacted with *p*-tolylacetylene (**19o**, 34 μL, 0.27 mmol), 0.2 M $CuSO_4$ (134 μL, 27 μmol) and 1 M sodium ascorbate (54 μL, 54 μmol) to yield **20o** (64 mg, 98%).

$[\alpha]_D$ +62.4 (*c* 1.09, $CHCl_3$); ^1H NMR (500 MHz, $CDCl_3$): δ 2.06, 2.07, 2.09, 2.19 (4 s, 12H, 4 $COCH_3$), 2.39 (s, 3H, PhCH_3), 3.94 (ddd, *J* = 2.4, 5.3, 9.0 Hz, 1H, H-5), 4.07 (dd, *J* = 2.4, 12.5 Hz, 1H, H-6a), 4.38 (dd, *J* = 5.4, 12.5 Hz, 1H, H-6b), 5.39 (t, *J* = 8.9 Hz, 1H, H-4), 5.98 (dd, *J* = 3.7, 8.7 Hz, 1H, H-3), 6.01 (dd, *J* = 2.6, 3.7 Hz, 1H, H-2), 6.05 (d, *J* = 2.4 Hz, 1H, H-1), 7.26, 7.74 (AA′, BB′ of AA′BB′, *J* = 7.9 Hz, 4H, C_6H_4), 7.90 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.6, 20.7, 20.8 (4C, 4 $COCH_3$), 21.3 (PhCH_3), 61.6 (C-6), 66.1 (C-4), 68.3 (C-2), 68.8 (C-3), 72.1 (C-5), 83.6 (C-1), 119.3 (C_2N_3H–C5), 125.8, 126.8, 129.6, 138.6 (6C, C_6H_4), 148.4 (C_2N_3H–C4), 169.3, 169.7, 169.7, 170.5 (4 $COCH_3$); ESI-MS Calcd for $C_{23}H_{27}N_3NaO_9$ $[M+Na]^+$: 512.16, Found: 512.15.

4.1.6.14. 1-(2,3,4,6-Tetra-*O*-acetyl-α-D-mannopyranosyl)-4-(3-chlorophenyl)-1,2,3-triazole (20p). Following general procedure C, **18** (50 mg, 0.13 mmol) was reacted with 3-chloro-1-ethinylbenzene (**19p**, 33 μL, 0.27 mmol), 0.2 M $CuSO_4$ (134 μL, 27 μmol) and 1 M sodium ascorbate (54 μL, 54 μmol) to yield **20p** (59 mg, 86%).

$[\alpha]_D$ +56.3 (*c* 1.03, $CHCl_3$); ^1H NMR (500 MHz, $CDCl_3$): δ 2.07, 2.08, 2.09, 2.19 (4 s, 12H, 4 $COCH_3$), 3.95 (m, 1H, H-5), 4.08 (dd, *J* = 2.1, 12.5 Hz, 1H, H-6a), 4.41 (dd, *J* = 5.4, 12.5 Hz, 1H, H-6b), 5.38 (t, *J* = 8.7 Hz, 1H, H-4), 5.95 (dd, *J* = 3.6, 8.6 Hz, 1H, H-3), 6.00 (m, 1H, H-2), 6.07 (d, *J* = 2.4 Hz, 1H, H-1), 7.34–7.40 (m, 2H, C_6H_4–H5, H6), 7.75 (d, *J* = 7.4 Hz, 1H, C_6H_4–H4), 7.85 (s, 1H, C_6H_4-H2), 7.98 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.6, 20.7, 20.7, 20.8 (4 $COCH_3$), 61.5 (C-6), 66.1 (C-4), 68.2 (C-2), 68.7 (C-3), 72.4 (C-5), 83.5 (C-1), 120.1 (C_2N_3H–C5), 124.0 (C_6H_4–C4), 126.0 (C_6H_4–C2), 128.7, 130.3 (C_6H_4–C5, C6), 131.5 (C_6H_4–C3), 134.9 (C_6H_4–C1), 147.2 (C_2N_3H–C4), 169.3, 169.6, 169.7, 170.5 (4 $COCH_3$); ESI-MS Calcd for $C_{22}H_{24}ClN_3NaO_9$ $[M+Na]^+$: 532.11, Found: 532.13.

4.1.6.15. 1-(2,3,4,6-Tetra-*O*-acetyl-α-D-mannopyranosyl)-4-(4-trifluoromethylphenyl)-1,2,3-triazole (20q). Following general procedure C, **18** (50 mg, 0.13 mmol) was reacted with 1-ethinyl-4-trifluoromethylbenzene (**19q**, 44 μL, 0.27 mmol), 0.2 M $CuSO_4$ (134 μL, 27 μmol) and 1 M sodium ascorbate (54 μL, 54 μmol) to yield **20q** (62 mg, 85%).

$[\alpha]_D$ +54.5 (*c* 0.95, $CHCl_3$); ^1H NMR (500 MHz, $CDCl_3$): δ 2.08, 2.09, 2.19 (3 s, 12H, 4 $COCH_3$), 3.97 (ddd, *J* = 2.5, 5.4, 8.7 Hz, 1H, H-5), 4.09 (dd, *J* = 2.5, 12.5 Hz, 1H, H-6a), 4.41 (dd, *J* = 5.5, 12.5 Hz, 1H, H-6b), 5.39 (t, *J* = 8.8 Hz, 1H, H-4), 5.95 (dd, *J* = 3.7, 8.7 Hz, 1H, H-3), 6.00 (dd, *J* = 3.1, 3.4 Hz, 1H, H-2), 6.09 (d, *J* = 2.8 Hz, 1H, H-1), 7.71, 7.98 (AA′, BB′ of AA′BB′, *J* = 8.1 Hz, 4H, C_6H_4), 8.05 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.60, 20.69, 20.70, 20.73 (4 $COCH_3$), 61.4 (C-6), 66.0 (C-4), 68.2 (C-2), 68.7 (C-3), 72.5 (C-5), 83.5 (C-1), 120.6 (C_2N_3H–C5), 124.0 (q, *J* = 272 Hz, CF_3), 126.0 (q, *J* = 3.8 Hz, 2C, C_6H_4–C3, C5), 126.1 (2C, C_6H_4–C2, C6), 130.5 (d, *J* = 32.6 Hz, C_6H_4–C4), 133.1 (C_6H_4–C1), 147.0 (C_2N_3H–C4), 169.3, 169.6, 169.7, 170.5 (4 $COCH_3$); ESI-MS Calcd for $C_{23}H_{24}F_3N_3NaO_9$ $[M+Na]^+$: 566.14, Found: 566.10.

4.1.6.16. 1-(2,3,4,6-Tetra-*O*-acetyl-α-D-mannopyranosyl)-4-(3-pyridyl)-1,2,3-triazole (20r). Following general procedure C, **18** (50 mg, 0.13 mmol) was reacted with 3-ethinylpyridine (**19r**, 27.6 mg, 0.27 mmol), 0.2 M $CuSO_4$ (134 μL, 27 μmol) and 1 M sodium ascorbate (54 μL, 54 μmol). Then the reaction mixture was diluted with DCM (20 mL) and extracted with 25 mM aq EDTA (10 mL). The organic layer was dried (Na_2SO_4), concentrated and the residue was purified by MPLC on silica (petrol ether/EtOAc) to yield **20r** (61 mg, 96%).

$[\alpha]_D$ +56.0 (*c* 0.70, $CHCl_3$); ^1H NMR (500 MHz, $CDCl_3$): δ 2.06, 2.06, 2.07, 2.17 (4 s, 12H, 4 $COCH_3$), 3.95 (ddd, *J* = 2.6, 5.5, 8.8 Hz, 1H, H-5), 4.07 (dd, *J* = 2.6, 12.5 Hz, 1H, H-6a), 4.40 (dd, *J* = 5.6, 12.5 Hz, 1H, H-6b), 5.38 (t, *J* = 8.8 Hz, 1H, H-4), 5.92 (dd, *J* = 3.7, 8.8 Hz, 1H, H-3), 6.00 (dd, *J* = 3.2, 3.5 Hz, 1H, H-2), 6.11 (d, *J* = 2.9 Hz, 1H, H-1), 7.45 (dd, *J* = 4.9, 7.2 Hz, 1H, C_5H_4N-H5), 8.11 (s, 1H, C_2N_3H), 8.28 (d, *J* = 7.9 Hz, 1H, C_5H_4N-H6), 8.62 (m, 1H, C_5H_4N-H4), 9.06 (br s, 1H, C_5H_4N–H2); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.56, 20.66, 20.68, 20.70 (4 $COCH_3$), 61.4 (C-6), 66.0 (C-4), 68.1 (C-2), 68.6 (C-3), 72.5 (C-5), 83.6 (C-1), 120.4 (C_2N_3H–C5), 124.1

(C_5H_4N–C5), 126.3 (C_5H_4N–C1), 133.7 (C_5H_4N–C6), 145.0 (C_2N_3H–C4), 146.5 (C_5H_4N–C2), 148.9 (C_5H_4N–C4), 169.3, 169.6, 169.6, 170.5 (4 $COCH_3$); ESI-MS Calcd for $C_{21}H_{24}N_4NaO_9$ $[M+Na]^+$: 477.16, Found: 477.08.

4.1.6.17. 1-(2,3,4,6-Tetra-*O*-acetyl-α-D-mannopyranosyl)-4-phenoxymethyl-1,2,3-triazole (20s). Following general procedure C, **18** (50 mg, 0.13 mmol) was reacted with phenylpropargyl ether (**19s**, 34 μL, 0.27 mmol), 0.2 M $CuSO_4$ (134 μL, 27 μmol) and 1 M sodium ascorbate (54 μL, 54 μmol) to yield **20s** (58 mg, 85%).

$[\alpha]_D$ +38.8 (*c* 0.61, $CHCl_3$); ^1H NMR (500 MHz, $CDCl_3$): δ 2.06, 2.07, 2.09, 2.18 (4 s, 12H, 4 $COCH_3$), 3.90 (m, 1H, H-5), 4.05 (dd, *J* = 2.0, 12.5 Hz, 1H, H-6a), 4.37 (dd, *J* = 5.4, 12.5 Hz, 1H, H-6b), 5.26 (s, 2H, CH_2OPh), 5.37 (t, *J* = 8.9 Hz, 1H, H-4), 5.93 (dd, *J* = 3.6, 8.8 Hz, 1H, H-3), 5.97 (dd, *J* = 2.7, 3.3 Hz, 1H, H-2), 6.00 (d, *J* = 2.3 Hz, 1H, H-1), 6.98–7.00, 7.30–7.32 (m, 5H, C_6H_5), 7.81 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.59, 20.67, 20.69, 20.73 (4 $COCH_3$), 61.5 (C-6), 61.8 (CH_2OPh), 66.0 (C-4), 68.2 (C-2), 68.7 (C-3), 72.2 (C-5), 83.6 (C-1), 114.7 (2C, C_6H_5–C2, C6), 121.4 (C_6H_5–C4), 123.0 (C_2N_3H–C5), 129.6 (2C, C_6H_5–C3, C5), 145.2 (C_2N_3H–C4), 158.0 (C_6H_5–C1), 169.3, 169.6, 169.7, 170.5 (4 $COCH_3$); ESI-MS Calcd for $C_{23}H_{27}N_3NaO_{10}$ $[M+Na]^+$: 528.16, Found: 528.14.

4.1.6.18. 1-(2,3,4,6-Tetra-*O*-acetyl-α-D-mannopyranosyl)methyl-4-phenyl-1,2,3-triazole (25n). Following general procedure C, **24** (40 mg, 0.10 mmol) was reacted with phenylacetylene (**19n**, 23 μL, 0.21 mmol), 0.2 M $CuSO_4$ (103 μL, 21 μmol) and 1 M sodium ascorbate (41 μL, 41 μmol) to yield **25n** (47 mg, 93%).

$[\alpha]_D$ -1.76 (*c* 1.50, $CHCl_3$); ^1H NMR (500 MHz, $CDCl_3$): δ 1.83, 2.08, 2.11, 2.12 (4 s, 12H, 4 $COCH_3$), 4.00 (dd, *J* = 3.9, 12.2 Hz, 1H, H-6a), 4.16 (dt, *J* = 3.5, 9.0 Hz, 1H, H-5), 4.39 (dt, *J* = 2.6, 8.6 Hz, 1H, H-1), 4.50 (dd, *J* = 8.8, 14.3 Hz, 1H, H-1′a), 4.59 (dd, *J* = 9.2, 12.1 Hz, 1H, H-6b), 4.66 (dd, *J* = 2.4, 14.3 Hz, 1H, H-1′b), 4.94 (dd, *J* = 3.0, 4.9 Hz, 1H, H-4), 5.06 (dd, *J* = 3.2, 8.4 Hz, 1H, H-2), 5.35 (t, *J* = 4.0 Hz, 1H, H-3), 7.32 (t, *J* = 7.4 Hz, 1H, C_6H_5-H4), 7.41 (t, *J* = 7.4 Hz, 2H, C_6H_5-H3, H5), 7.83 (d, *J* = 8.0 Hz, 2H, C_6H_5-H2, H6), 7.92 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.4, 20.6, 20.7, 20.8 (4 $COCH_3$), 50.3 (C-1′), 60.0 (C-6), 66.8 (C-2), 66.9 (C-3), 68.0 (C-4), 68.4 (C-1), 73.4 (C-5), 120.9 (C_2N_3H–C5), 125.6, 128.2, 128.8, 130.5 (6C, C_6H_5), 148.0 (C_2N_3H–C4), 169.1, 169.4, 169.6, 170.4 (4 $COCH_3$); ESI-MS Calcd for $C_{23}H_{28}N_3O_9$ $[M+H]^+$: 490.18, Found: 490.17.

4.1.6.19. 1-(2,3,4,6-Tetra-*O*-acetyl-α-D-mannopyranosyl)methyl-4-(4-methylphenyl)-1,2,3-triazole (25o). Following general procedure C, **24** (40 mg, 0.10 mmol) was reacted with *p*-tolylacetylene (**19o**, 26 μL, 0.21 mmol), 0.2 M $CuSO_4$ (103 μL, 21 μmol) and 1 M sodium ascorbate (41 μL, 41 μmol) to yield **25o** (50 mg, 97%).

$[\alpha]_D$ -1.57 (*c* 1.26, $CHCl_3$); ^1H NMR (500 MHz, $CDCl_3$): δ 1.83, 2.07, 2.10, 2.11 (4 s, 12H, 4 $COCH_3$), 2.36 (s, 3H, PhCH_3), 3.99 (dd, *J* = 4.0, 12.2 Hz, 1H, H-6a), 4.15 (dt, *J* = 3.5, 9.0 Hz, 1H, H-5), 4.38 (dt, *J* = 2.7, 8.6 Hz, 1H, H-1), 4.50 (dd, *J* = 8.8, 14.4 Hz, 1H, H-1′a), 4.58 (dd, *J* = 9.1, 12.2 Hz, 1H, H-6b), 4.64 (dd, *J* = 2.8, 14.4 Hz, 1H, H-1′b), 4.94 (dd, *J* = 3.1, 5.0 Hz, 1H, H-4), 5.06 (dd, *J* = 3.3, 8.3 Hz, 1H, H-2), 5.34 (dd, *J* = 3.5, 4.8 Hz, 1H, H-3), 7.21, 7.70 (AA′, BB′ of AA′BB′, *J* = 8.0 Hz, 4H, C_6H_4), 7.87 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.41, 20.62, 20.64, 20.74 (4 $COCH_3$), 21.2 (PhCH_3), 50.2 (C-1′), 60.0 (C-6), 66.8 (C-2), 66.9 (C-3), 68.0 (C-4), 68.5 (C-1), 73.3 (C-5), 120.5 (C_2N_3H–C5), 125.5, 127.6, 129.5, 138.0 (6C, C_6H_4), 147.8 (C_2N_3H–C4), 169.1, 169.4, 169.6, 170.4 (4 $COCH_3$); ESI-MS Calcd for $C_{24}H_{30}N_3O_9$ $[M+H]^+$: 504.20, Found: 504.20.

4.1.6.20. 1-(2,3,4,6-Tetra-*O*-acetyl-α-D-mannopyranosyl)methyl-4-(3-chlorophenyl)-1,2,3-triazole (25p). Following general procedure C, **24** (40 mg, 0.10 mmol) was reacted with 3-chloro-1-ethinylbenzene (**19p**, 25 μL, 0.21 mmol), 0.2 M $CuSO_4$ (103 μL, 21 μmol) and 1 M sodium ascorbate (41 μL, 41 μmol) to yield **25p** (51 mg, 94%).

$[\alpha]_D$ -2.55 (*c* 1.27, $CHCl_3$); ^1H NMR (500 MHz, $CDCl_3$): δ 1.84, 2.07, 2.10, 2.11 (4 s, 12H, 4 $COCH_3$), 3.99 (dd, *J* = 4.0, 12.2 Hz, 1H, H-6a), 4.15 (dt, *J* = 3.4, 9.0 Hz, 1H, H-5), 4.38 (dt, *J* = 2.7, 8.6 Hz, 1H, H-1), 4.51 (dd, *J* = 8.7, 14.4 Hz, 1H, H-1′a), 4.60 (dd, *J* = 9.2, 12.2 Hz, 1H, H-6b), 4.65 (dd, *J* = 2.7, 14.4 Hz, 1H, H-1′b), 4.93 (dd, *J* = 2.9, 4.9 Hz, 1H, H-4), 5.03 (dd, *J* = 3.3, 8.5 Hz, 1H, H-2), 5.34 (dd, *J* = 3.6, 4.6 Hz, 1H, H-3), 7.27 (m, 1H, C_6H_4-H6), 7.33 (t, *J* = 7.8 Hz, 1H, C_6H_4–H5), 7.72 (d, *J* = 7.7 Hz, 1H, C_6H_4–H4), 7.81 (t, *J* = 1.7 Hz, 1H, C_6H_4-H2), 7.94 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.41, 20.60, 20.64, 20.73 (4 $COCH_3$), 50.4 (C-1′), 59.9 (C-6), 66.7 (C-2), 66.8 (C-3), 68.0 (C-4), 68.3 (C-1), 73.5 (C-5), 121.3 (C_2N_3H–C5), 123.6, 125.6, 128.1, 130.1, 132.3, 134.8 (C_6H_4), 146.5 (C_2N_3H–C4), 169.1, 169.3, 169.5, 170.3 (4 $COCH_3$); ESI-MS Calcd for $C_{23}H_{27}ClN_3O_9$ $[M+H]^+$: 524.14, Found: 524.04.

4.1.6.21. 1-(2,3,4,6-Tetra-*O*-acetyl-α-D-mannopyranosyl)methyl-4-(4-trifluoromethylphenyl)-1,2,3-triazole (25q). Following general procedure C, **24** (40 mg, 0.10 mmol) was reacted with 1-ethinyl-4-trifluoromethylbenzene (**19q**, 34 μL, 0.21 mmol), 0.2 M $CuSO_4$ (103 μL, 21 μmol) and 1 M sodium ascorbate (41 μL, 41 μmol) to yield **25q** (56 mg, 98%).

$[\alpha]_D$ +0.47 (*c* 1.19, $CHCl_3$); ^1H NMR (500 MHz, $CDCl_3$): δ 1.82, 2.08, 2.10, 2.12 (4 s, 12H, 4 $COCH_3$), 3.97 (dd, *J* = 4.0, 12.2 Hz, 1H, H-6a), 4.15 (dt, *J* = 3.4, 9.0 Hz, 1H, H-5), 4.40 (dt, *J* = 2.5, 8.6 Hz, 1H, H-1), 4.53 (dd, *J* = 8.7, 14.4 Hz, 1H, H-1′a), 4.63 (dd, *J* = 9.2, 12.2 Hz, 1H, H-6b), 4.67 (dd, *J* = 2.7, 14.4 Hz, 1H, H-1′b), 4.93 (dd, *J* = 2.9, 4.7 Hz, 1H, H-4), 5.04 (dd, *J* = 3.2, 8.6 Hz, 1H, H-2), 5.34 (t, *J* = 4.0 Hz, 1H, H-3), 7.66, 7.95 (AA′, BB′ of AA′BB′, *J* = 8.2 Hz, 4H, C_6H_4), 8.00 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.42, 20.63, 20.67, 20.76 (4 $COCH_3$), 50.4 (C-1′), 59.9 (C-6), 66.7 (C-2), 66.8 (C-3), 68.0 (C-4), 68.2 (C-1), 73.5 (C-5), 121.7 (C_2N_3H–C5), 124.0 (q, *J* = 272 Hz, CF_3), 125.7 (2C, C_6H_4–C2, C6), 125.8 (q, *J* = 3.8 Hz, 2C, C_6H_4–C3, C5), 129.9 (q, *J* = 32.5 Hz, C_6H_4–C4), 133.9 (C_6H_4–C1), 146.4 (C_2N_3H–C4), 169.1, 169.4, 169.6, 170.3 (4 $COCH_3$); ESI-MS Calcd for $C_{24}H_{27}F_3N_3O_9$ $[M+H]^+$: 558.17, Found: 558.22.

4.1.6.22. 1-(2,3,4,6-Tetra-*O*-acetyl-α-D-mannopyranosyl)methyl-4-(3-pyridyl)-1,2,3-triazole (25r). Following general procedure C, **24** (40 mg, 0.10 mmol) was reacted with 3-ethinylpyridine (**19r**, 21.2 mg, 0.206 mmol), 0.2 M $CuSO_4$ (103 μL, 21 μmol) and 1 M sodium ascorbate (41 μL, 41 μmol) to yield **25r** (50 mg, 98%).

$[\alpha]_D$ -0.08 (*c* 1.04, $CHCl_3$); ^1H NMR (500 MHz, $CDCl_3$): δ 1.83, 2.07, 2.09, 2.10 (4 s, 12H, 4 $COCH_3$), 3.97 (dd, *J* = 4.0, 12.2 Hz, 1H, H-6a), 4.14 (dt, *J* = 3.4, 8.9 Hz, 1H, H-5), 4.39 (dt, *J* = 2.6, 8.6 Hz, 1H, H-1), 4.54 (dd, *J* = 8.6, 14.4 Hz, 1H, H-1′a), 4.60 (dd, *J* = 9.1, 12.2 Hz, 1H, H-6b), 4.66 (dd, *J* = 2.7, 14.4 Hz, 1H, H-1′b), 4.92 (dd, *J* = 2.9, 4.8 Hz, 1H, H-4), 5.02 (dd, *J* = 3.3, 8.5 Hz, 1H, H-2), 5.33 (m, 1H, H-3), 7.37 (m, 1H, C_5H_4N-H5), 8.02 (s, 1H, C_2N_3H), 8.21 (d, *J* = 7.9 Hz, 1H, C_5H_4N-H6), 8.56 (s, 1H, C_5H_4N–H2), 9.00 (m, 1H, C_5H_4N-H4); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.44, 20.61, 20.65, 20.74 (4 $COCH_3$), 50.4 (C-1′), 59.9 (C-6), 66.6 (C-2), 66.8 (C-3), 67.9 (C-4), 68.3 (C-1), 73.5 (C-5), 121.3 (C_2N_3H–C5), 123.8 (C_5H_4N–C5), 126.8 (C_5H_4N–C1), 133.0 (C_5H_4N–C6), 144.6 (C_2N_3H–C4), 146.7 (C_5H_4N–C2), 149.0 (C_5H_4N–C4), 169.1, 169.3, 169.5, 170.3 (4 $COCH_3$); ESI-MS Calcd for $C_{22}H_{27}N_4O_9$ $[M+H]^+$: 491.18, Found: 491.17.

4.1.6.23. 1-(2,3,4,6-Tetra-*O*-acetyl-α-D-mannopyranosyl)methyl-4-phenoxymethyl-1,2,3-triazole (25s). Following general procedure C, **24** (40 mg, 0.10 mmol) was reacted with phenylpropargyl ether (**19s**, 26 μL, 0.21 mmol), 0.2 M $CuSO_4$ (103 μL, 21 μmol) and 1 M sodium ascorbate (41 μL, 41 μmol) to yield **25s** (51 mg, 96%).

$[\alpha]_D$ +2.34 (*c* 1.03, $CHCl_3$); ^1H NMR (500 MHz, $CDCl_3$): δ 2.00, 2.08, 2.12, 2.13 (4 s, 12H, 4 $COCH_3$), 4.04 (dd, *J* = 4.1, 12.1 Hz, 1H, H-6a), 4.15 (m, 1H, H-5), 4.37 (dt, *J* = 2.4, 8.6 Hz, 1H, H-1), 4.50 (dd, *J* = 8.7, 14.4 Hz, 1H, H-1′a), 4.55 (dd, *J* = 9.0, 12.1 Hz, 1H, H-6b), 4.63 (dd, *J* = 2.5, 14.4 Hz, 1H, H-1′b), 4.96 (dd, *J* = 3.2, 4.9 Hz, 1H, H-4), 5.04 (dd, *J* = 3.3, 8.3 Hz, 1H, H-2), 5.34 (m, 1H, H-3), 6.96–6.99, 7.29–7.31 (m, 5H, C_6H_5), 7.80 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 20.49, 20.62, 20.66, 20.74 (4 $COCH_3$), 50.2 (C-1′), 60.0 (C-6), 61.9 (CH_2OPh), 66.7 (C-2), 66.9 (C-3), 67.8 (C-4), 68.5 (C-1), 73.3 (C-5), 114.6, 121.2 (3C, C_6H_5), 123.8 (C_2N_3H–C5), 129.5 (2C, C_6H_5), 144.3 (C_2N_3H–C4), 158.1 (C_6H_4–C1), 169.1, 169.4, 169.6, 170.4 (4 $COCH_3$); ESI-MS Calcd for $C_{24}H_{30}N_3O_9$ $[M+H]^+$: 520.19, Found: 520.16.

4.1.7. Synthesis of mannosyl triazoles

4.1.7.1. Sodium 4-[4-((α-D-mannopyranosyloxy)methyl)-1*H*-1,2,3-triazol-1-yl]-benzoate (5a). To a solution of **12a** (48.0 mg, 85.2 μmol) in MeOH (4 mL) was added freshly prepared 1 M NaOMe in MeOH (0.4 mL) under argon. The mixture was stirred at rt for 3 h and then evaporated to dryness. The remains were dissolved in H_2O/dioxane (1:1, 5 mL) and treated with 1 M aq. NaOH (0.5 mL) for 16 h. The solution was concentrated and the residue purified by MPLC on RP-18 (H_2O/MeOH) and P2 size exclusion chromatography to give **5a** (31.2 mg, 91%) as white powder after a final lyophilization from water.

$[\alpha]_D$ +41.7 (*c* 0.60, H_2O); IR (KBr) 3413 (vs b, OH), 1607 (vs, CO) cm^{-1}; ^1H NMR (500 MHz, D_2O): δ 3.60–3.66 (m, 2H, H-4, H-5), 3.72 (dd, *J* = 5.5, 12.1 Hz, 1H, H-6a), 3.78 (dd, *J* = 3.4, 9.4 Hz, 1H, H-3), 3.82 (dd, *J* = 1.4, 12.2 Hz, 1H, H-6b), 3.93 (dd, *J* = 1.7, 3.4 Hz, 1H, H-2), 4.71 (m, 1H, H-1′a), 4.82 (m, 1H, H-1′b), 4.97 (s, 1H, H-1), 7.68 (m, 2H, C_6H_4), 7.94 (m, 2H, C_6H_4), 8.41 (m, 1H, C_2N_3H); ^{13}C NMR (125 MHz, D_2O): δ 59.7 (C-1′), 60.8 (C-6), 66.6 (C-4), 69.9 (C-2), 70.4 (C-3), 72.9 (C-5), 99.5 (C-1), 120.2 (2C, C_6H_4), 123.3 (C_2N_3H–C5), 130.3, 137.0, 137.7 (4C, C_6H_4), 144.2 (C_2N_3H–C4), 174.1 (CO); HR-MS Calcd for $C_{16}H_{19}N_3NaO_8$ $[M+H]^+$: 404.1070, Found: 404.1071.

4.1.7.2. Sodium 3-[4-((α-D-mannopyranosyloxy)methyl)-1*H*-1,2,3-triazol-1-yl]-benzoate (5b). According to the procedure described for **5a**, **12b** (67.0 mg, 0.116 mmol) was subsequently treated with 1 M methanolic NaOMe (0.5 mL) in MeOH (5 mL) and 1 M aq. NaOH (0.5 mL) in H_2O/dioxane (1:1, 6 mL) to yield **5b** (37.7 mg, 81%).

$[\alpha]_D$ +44.4 (*c* 0.89, H_2O); IR (KBr) 3401 (vs b, OH), 1610 (vs, CO) cm^{-1}; ^1H NMR (500 MHz, D_2O): δ 3.61–3.66 (m, 2H, H-4, H-5), 3.72 (dd, *J* = 5.4, 12.2 Hz, 1H, H-6a), 3.78 (dd, *J* = 3.5, 9.5 Hz, 1H, H-3), 3.82 (dd, *J* = 1.5, 12.2 Hz, 1H, H-6b), 3.92 (dd, *J* = 1.7, 3.3 Hz, 1H, H-2), 4.69 (A of AB, *J* = 12.8 Hz, 1H, H-1′a), 4.80 (m, 1H, H-1′b), 4.96 (s, 1H, H-1), 7.51 (m, 1H, C_6H_4–H5), 7.69 (d, *J* = 7.0 Hz, 1H, C_6H_4–H4), 7.87 (d, *J* = 7.7 Hz, 1H, C_6H_4-H6), 8.01 (s, 1H, C_6H_4-H2), 8.36 (m, 1H, C_2N_3H); ^{13}C NMR (125 MHz, D_2O): δ 59.7 (C-1′), 60.8 (C-6), 66.7 (C-4), 69.9 (C-2), 70.4 (C-3), 72.9 (C-5), 99.5 (C-1), 120.8 (C_6H_4), 123.0 (C_2N_3H–C5), 123.3, 129.5, 129.8, 135.9, 138.1 (C_6H_4), 144.1 (C_2N_3H–C4), 173.6 (CO); HR-MS Calcd for $C_{16}H_{19}N_3NaO_8$ $[M+H]^+$: 404.1070, Found: 404.1068.

4.1.7.3. Sodium 5-[4-((α-D-mannopyranosyloxy)methyl)-1*H*-1,2,3-triazol-1-yl]-nicotinate (5c). According to the procedure described for **5a**, **12c** (41.0 mg, 72.6 μmol) was subsequently treated with 1 M methanolic NaOMe (0.4 mL) in MeOH (4 mL) and 1 M aq. NaOH (0.4 mL) in H_2O/dioxane (1:1, 4 mL) to yield **5c** (23.0 mg, 78%).

$[\alpha]_D$ +36.0 (*c* 0.69, H_2O); IR (KBr) 3413 (vs b, OH), 1616 (vs, CO) cm^{-1}; ^1H NMR (500 MHz, D_2O): δ 3.61–3.66 (m, 2H, H-4, H-5), 3.73 (dd, *J* = 5.4, 12.1 Hz, 1H, H-6a), 3.79 (dd, *J* = 3.5, 9.4 Hz, 1H, H-3), 3.83 (d, *J* = 12.0 Hz, 1H, H-6b), 3.94 (dd, *J* = 1.7, 3.2 Hz, 1H, H-2), 4.76, 4.87 (A, B of AB, *J* = 12.6 Hz, 1H, H-1′), 4.99 (d, *J* = 0.7 Hz, 1H, H-1), 8.47 (t, *J* = 2.0 Hz, 1H, C_5H_3N-H2), 8.56 (m, 1H, C_2N_3H), 8.96 (d, *J* = 2.1 Hz, 1H, C_5H_3N-H6), 8.97 (d, *J* = 1.4 Hz, 1H, C_5H_3N-H4); ^{13}C NMR (125 MHz, D_2O): δ 59.7 (C-1′), 60.8 (C-6), 66.7 (C-4), 69.9 (C-2), 70.4 (C-3), 73.0 (C-5), 99.5 (C-1), 123.6 (C_2N_3H–C5), 129.5, 133.1, 133.4, 142.6 (C_5H_3N), 144.6 (C_2N_3H–C4), 149.7 (C_5H_3N), 171.1 (CO); HR-MS Calcd for $C_{15}H_{17}N_4Na_2O_8$ $[M+Na]^+$: 427.0842, Found: 427.0844.

4.1.7.4. (1-Benzyl-1,2,3-triazol-4-yl)methyl α-D-mannopyranoside (5d). Following general procedure A, **14** (40 mg, 0.18 mmol) was reacted with benzyl azide (**15d**, 34 μL, 0.27 mmol), $CuSO_4$ (11 mg, 45 μmol) and sodium ascorbate (18 mg, 90 μmol) to yield **5d** (57 mg, 71%).

$[\alpha]_D$ +53.3 (*c* 1.03, MeOH); ^1H NMR (500 MHz, CD_3OD): δ 3.50 (m, 1H, H-5), 3.56 (t, *J* = 9.4 Hz, 1H, H-4), 3.60–3.68 (m, 2H, H-3, H-6a), 3.73 (m, 1H, H-2), 3.79 (dd, *J* = 1.7, 11.7 Hz, 1H, H-6b), 4.60, 4.75 (A, B of AB, *J* = 12.4 Hz, 2H, H-1′), 4.80 (d, *J* = 1.6 Hz, 1H, H-1), 5.56 (s, 2H, CH_2Ph), 7.24–7.41 (m, 5H, C_6H_5), 7.97 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 55.1 (CH_2Ph), 60.8 (C-1′), 63.1 (C-6), 68.7 (C-4), 72.1 (C-2), 72.6 (C-3), 75.1 (C-5), 100.9 (C-1), 125.5 (C_2N_3H–C5), 129.3, 129.8, 130.2, 136.9 (6C, C_6H_5), 145.9 (C_2N_3H–C4); HR-MS Calcd for $C_{16}H_{21}N_3NaO_6$ $[M+Na]^+$: 374.1328, Found: 374.1334.

4.1.7.5. [1-(4′-Aminophenyl)-1,2,3-triazol-4-yl]methyl α-D-mannopyranoside hydrochloride (5e). Following general procedure A, **14** (40 mg, 0.18 mmol) was reacted with 4-azidoaniline hydrochloride (**15e**, 46 mg, 0.27 mmol), $CuSO_4$ (11 mg, 45 μmol) and sodium ascorbate (18 mg, 90 μmol) to yield **5e** (19 mg, 27%).

$[\alpha]_D$ +55.2 (*c* 1.00, MeOH); ^1H NMR (500 MHz, CD_3OD): δ 3.55–3.60 (m, 2H, H-4, H-5), 3.64–3.72 (m, 2H, H-3, H-6a), 3.78 (m, 1H, H-2), 3.83 (m, 1H, H-6b), 4.68, 4.82 (A, B of AB, *J* = 12.4 Hz, 2H, H-1′), 4.86 (m, 1H, H-1), 6.78, 7.45 (AA′, BB′ of AA′BB′, *J* = 8.7 Hz, 4H, C_6H_4), 8.33 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 60.8 (C-1′), 63.2 (C-6), 68.8 (C-4), 72.2 (C-2), 72.6 (C-3), 75.2 (C-5), 100.9 (C-1), 116.2 (2C, C_6H_4), 123.3 (2C, C_6H_4), 123.7 (C_2N_3H–C5), 128.8 (C_6H_4–C1), 145.8 (C_2N_3H–C4), 150.8 (C_6H_4–C4); HR-MS Calcd for $C_{15}H_{21}N_4O_6$ $[M+H]^+$: 353.1461, Found: 353.1463.

4.1.7.6. (1-Adamantyl-1,2,3-triazol-4-yl)methyl α-D-mannopyranoside (5f). Following general procedure A, **14** (40 mg, 0.18 mmol) was reacted with 1-azidoadamantane (**15f**, 48 mg, 0.27 mmol), $CuSO_4$ (11 mg, 45 μmol) and sodium ascorbate (18 mg, 90 μmol) to yield **5f** (20 mg, 28%).

$[\alpha]_D$ +50.5 (*c* 1.04, MeOH); ^1H NMR (500 MHz, CD_3OD): δ 1.76–1.88 (m, 6H, Ad), 2.24 (s, 9H, Ad), 3.50–3.60 (m, 2H, H-4, H-5), 3.61–3.70 (m, 2H, H-3, H-6a), 3.75 (dd, *J* = 1.7, 3.3 Hz, 1H, H-2), 3.82 (m, 1H, H-6b), 4.60, 4.76 (A, B of AB, *J* = 12.3 Hz, 2H, H-1′), 4.80 (d, *J* = 1.4 Hz, 1H, H-1), 8.09 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 31.1, 37.1, 44.0 (10 C, Ad), 60.9 (C-1′), 63.2 (C-6), 68.8 (C-4), 72.2 (C-2), 72.6 (C-3), 75.1 (C-5), 100.9 (C-1), 122.2 (C_2N_3H–C5), 144.6 (C_2N_3H–C4); HR-MS Calcd for $C_{19}H_{29}N_3NaO_6$ $[M+Na]^+$: 418.1954, Found: 418.1951.

4.1.7.7. [1-(4′-Methoxybenzyl)-1,2,3-triazol-4-yl]methyl α-D-mannopyranoside (5g). Following general procedure A, **14** (50 mg, 0.23 mmol) was reacted with 4-methoxybenzylazide (**15g**,[38] 57 mg, 0.35 mmol), $CuSO_4$ (15 mg, 60 μmol) and sodium ascorbate (24 mg, 120 μmol) to yield **5g** (64 mg, 73%).

$[\alpha]_D$ +66.6 (*c* 1.01, MeOH); ^1H NMR (500 MHz, CD_3OD): δ 3.50 (m, 1H, H-5), 3.57 (t, *J* = 9.4 Hz, 1H, H-4), 3.61–3.69 (m, 2H, H-3, H-6a), 3.74 (m, 4H, H-2, OCH_3), 3.78 (dd, *J* = 1.7, 11.7 Hz, 1H, H-6b), 4.58, 4.73 (A, B of AB, *J* = 12.4 Hz, 2H, H-1′), 4.79 (m, 1H, H-1), 5.46 (s, 2H, CH_2Ar), 6.88, 7.25 (AA′, BB′ of AA′BB′, *J* = 8.6 Hz, 4H, C_6H_4), 7.91 (s, 1H, C_2N_3H);

^{13}C NMR (125 MHz, CD_3OD): δ 54.6 (CH_2Ar), 55.8 (OCH_3), 60.7 (C-1′), 63.0 (C-6), 68.6 (C-4), 72.0 (C-2), 72.5 (C-3), 75.0 (C-5), 100.8 (C-1), 115.4 (2C, C_6H_4), 125.1 (C_2N_3H–C5), 128.6 (2C, C_6H_4), 130.8 (C_6H_4–C1), 145.6 (C_2N_3H–C4), 161.4 (C_6H_4–C4); HR-MS Calcd for $C_{17}H_{23}N_3NaO_7$ [M+Na]$^+$: 404.1434, Found: 404.1431.

4.1.7.8. [1-(3′-Methoxybenzyl)-1,2,3-triazol-4-yl]methyl α-D-mannopyranoside (5h). Following general procedure A, **14** (50 mg, 0.23 mmol) was reacted with 3-methoxybenzylazide (**15h**,[39] 57 mg, 0.35 mmol), $CuSO_4$ (15 mg, 60 μmol) and sodium ascorbate (24 mg, 120 μmol) to yield **5h** (68 mg, 77%).

$[\alpha]_D$ +62.2 (*c* 1.00, MeOH); ^{1}H NMR (500 MHz, CD_3OD): δ 3.52 (m, 1H, H-5), 3.58 (t, *J* = 9.4 Hz, 1H, H-4), 3.61–3.70 (m, 2H, H-3, H-6a), 3.70–3.84 (m, 5H, H-2, H-6b, OCH_3), 4.60, 4.75 (A, B of AB, *J* = 12.3 Hz, 2H, H-1′), 4.81 (m, 1H, H-1), 5.52 (s, 2H, CH_2Ar), 6.85 (s, 3H, C_6H_4), 7.24 (t, *J* = 7.9 Hz, 1H, C_6H_4), 7.98 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 55.5 (CH_2Ar), 55.9 (OCH_3), 60.8 (C-1′), 63.0 (C-6), 68.7 (C-4), 72.1 (C-2), 72.6 (C-3), 75.1 (C-5), 100.9 (C-1), 114.9, 115.2, 121.4 (3C, C_6H_4), 125.6 (C_2N_3H–C5), 131.3 (C_6H_4), 138.2 (C_6H_4–C1), 145.8 (C_2N_3H–C4), 161.7 (C_6H_4–C3); HR-MS Calcd for $C_{17}H_{23}N_3NaO_7$ [M+Na]$^+$: 404.1434, Found: 404.1435.

4.1.7.9. [1-(4′-Nitrophenyl)-1,2,3-triazol-4-yl]methyl α-D-mannopyranoside (5i). Following general procedure A, **14** (40 mg, 0.18 mmol) was reacted with 1-azido-4-nitrobenzene (**15i**,[40] 44 mg, 0.27 mmol), $CuSO_4$ (11 mg, 45 μmol) and sodium ascorbate (18 mg, 90 μmol) to yield **5i** (31 mg, 44%).

$[\alpha]_D$ +50.4 (*c* 1.02, MeOH); ^{1}H NMR (500 MHz, CD_3OD): δ 3.56–3.60 (m, 2H, H-4, H-5), 3.64–3.72 (m, 2H, H-3, H-6a), 3.80 (dd, *J* = 1.7, 3.3 Hz, 1H, H-2), 3.84 (dd, *J* = 1.0, 11.7 Hz, 1H, H-66), 4.75 (A of AB, *J* = 12.5 Hz, 1H, H-1′a), 4.88–4.91 (m, 2H, H-1, H-1′b), 8.16 (m, 2H, C_6H_4), 8.44 (m, 2H, C_6H_4), 8.75 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 60.9 (C-1′), 63.2 (C-6), 68.8 (C-4), 72.1 (C-2), 72.6 (C-3), 75.3 (C-5), 101.2 (C-1), 122.0 (2C, C_6H_4), 123.9 (C_2N_3H–C5), 126.7 (2C, C_6H_4), 142.7 (C_6H_4–C1), 147.1 (C_2N_3H–C4), 148.9 (C_6H_4–C4); HR-MS Calcd for $C_{15}H_{18}N_4NaO_8$ [M+Na]$^+$: 405.1022, Found: 405.1020.

4.1.7.10. [1-(Pyridin-4′-yl)-1,2,3-triazol-4-yl]methyl α-D-mannopyranoside (5j). Prepared from **16j** (102 mg, 0.20 mmol) according to general procedure D. Yield: 58 mg, 85%.

$[\alpha]_D$ +70.3 (*c* 1.00, MeOH); ^{1}H NMR (500 MHz, CD_3OD): δ 3.53–3.63 (m, 2H, H-4, H-5), 3.64–3.74 (m, 2H, H-3, H-6a), 3.77–3.87 (m, 2H, H-2, H-6b), 4.73 (A of AB, *J* = 12.6 Hz, 1H, H-1′a), 4.84–4.92 (m, 2H, H-1, H-1′b), 7.96, 8.67 (m, 4H, C_5H_4N), 8.77 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 60.9 (C-1′), 63.1 (C-6), 68.7 (C-4), 72.1 (C-2), 72.6 (C-3), 75.2 (C-5), 101.2 (C-1), 115.5 (2C, C_5H_4N), 123.5 (C_2N_3H–C5), 145.2 (C_2N_3H–C4), 147.2 (C_5H_4N–C1), 152.5 (2C, C_5H_4N); HR-MS Calcd for $C_{14}H_{19}N_4O_6$ [M+H]$^+$: 339.1305, Found: 339.1302.

4.1.7.11. [1-(4′-Fluorophenyl)-1,2,3-triazol-4-yl]methyl α-D-mannopyranoside (5k). Prepared from **16k** (106 mg, 0.20 mmol) according to general procedure D. Yield: 56 mg, 78%.

$[\alpha]_D$ +78.5 (*c* 1.00, MeOH); ^{1}H NMR (500 MHz, CD_3OD): δ 3.60–3.69 (m, 2H, H-4, H-5), 3.71–3.78 (m, 2H, H-3, H-6a), 3.86 (dd, *J* = 1.7, 3.4 Hz, 1H, H-2), 3.89 (dd, *J* = 1.8, 11.8 Hz, 1H, H-6b), 4.77, 4.91 (A, B of AB, *J* = 12.5 Hz, 2H, H-1′), 4.94 (d, *J* = 1.6 Hz, 1H, H-1), 7.34, 7.89 (m, 4H, C_6H_4), 8.56 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 60.9 (C-1′), 63.2 (C-6), 68.8 (C-4), 72.2 (C-2), 72.6 (C-3), 75.2 (C-5), 101.1 (C-1), 117.8 (d, *J* = 23.8 Hz, 2C, C_6H_4), 124.0 (2C, C_6H_4), 124.0 (C_2N_3H–C5), 134.9 (d, *J* = 3.8 Hz, C_6H_4–C1), 146.5 (C_2N_3H–C4), 164.1 (d, *J* = 246.3 Hz, C_6H_4–C4); HR-MS Calcd for $C_{15}H_{18}FN_3NaO_6$ [M+Na]$^+$: 378.1077, Found: 378.1079.

4.1.7.12. [1-(3′-Fluorophenyl)-1,2,3-triazol-4-yl]methyl α-D-mannopyranoside (5l). Prepared from **16l** (105 mg, 0.20 mmol) according to general procedure D. Yield: 58 mg, 81%.

$[\alpha]_D$ +73.8 (*c* 1.00, MeOH); ^{1}H NMR (500 MHz, CD_3OD): δ 3.51–3.62 (m, 2H, H-4, H-5), 3.62–3.73 (m, 2H, H-3, H-6a), 3.78 (dd, *J* = 1.6, 3.1 Hz, 1H, H-2), 3.84 (m, 1H, H-6b), 4.71 (A of AB, *J* = 12.4 Hz, 1H, H-1′a), 4.82–4.89 (m, 2H, H-1, H-1′b), 7.23 (m, 1H, C_6H_4), 7.57 (m, 1H, C_6H_4), 7.64–7.74 (m, 2H, C_6H_4), 8.61 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 60.8 (C-1′), 63.1 (C-6), 68.8 (C-4), 72.1 (C-2), 72.6 (C-3), 75.2 (C-5), 101.1 (C-1), 109.1 (d, *J* = 26.3 Hz, C_6H_4), 116.9 (d, *J* = 21.3 Hz, C_6H_4), 117.3 (d, *J* = 3.8 Hz, C_6H_4), 123.8 (C_2N_3H–C5), 132.9 (d, *J* = 10.0 Hz, C_6H_4), 139.7 (d, *J* = 10.0 Hz, C_6H_4–C1), 146.6 (C_2N_3H–C4), 161.7 (d, *J* = 245.0 Hz, C_6H_4–C3); HR-MS Calcd for $C_{15}H_{18}FN_3NaO_6$ [M+Na]$^+$: 378.1077, Found: 378.1081.

4.1.7.13. [1-(4′-Methoxyphenyl)-1,2,3-triazol-4-yl]methyl α-D-mannopyranoside (5m). Prepared from **16m** (113 mg, 0.21 mmol) according to general procedure D. Yield: 58 mg, 75%.

$[\alpha]_D$ +37.4 (*c* 1.01, MeOH); ^{1}H NMR (500 MHz, CD_3OD): δ 3.54–3.64 (m, 2H, H-4, H-5), 3.65–3.74 (m, 2H, H-3, H-6a), 3.80 (dd, *J* = 1.6, 3.2 Hz, 1H, H-2), 3.81–3.87 (m, 4H, H-6b, OCH_3), 4.69, 4.83 (A, B of AB, *J* = 12.4 Hz, 2H, H-1′), 4.87 (m, 1H, H-1), 7.05, 7.68 (AA′, BB′ of AA′BB′, *J* = 9.0 Hz, 4H, C_6H_4), 8.42 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 56.3 (OCH_3), 60.8 (C-1′), 63.1 (C-6), 68.8 (C-4), 72.2 (C-2), 72.6 (C-3), 75.2 (C-5), 101.0 (C-1), 116.0 (2C, C_6H_4), 123.4 (2C, C_6H_4), 123.8 (C_2N_3H–C5), 131.7 (C_6H_4–C1), 146.1 (C_2N_3H–C4), 161.7 (C_6H_4–C4); HR-MS Calcd for $C_{16}H_{21}N_3NaO_7$ [M+Na]$^+$: 390.1277, Found: 390.1279.

4.1.7.14. [1-(3′-Methoxybenzyl)-1,2,3-triazol-4-yl]ethyl α-D-mannopyranoside (6h). Prepared from **17h** (53 mg, 94 μmol) according to general procedure D. Yield: 30 mg, 81%.

$[\alpha]_D$ +45.9 (*c* 1.00, MeOH); ^{1}H NMR (500 MHz, CD_3OD): δ 2.99 (t, *J* = 6.6 Hz, 2H, H-2′), 3.41 (m, 1H, H-5), 3.57–3.75 (m, 4H, H-3, H-4, H-6a, H-1′a), 3.71 (dd, *J* = 1.7, 3.1 Hz, 1H, H-2), 3.78–3.83 (m, 4H, H-6b, OCH_3), 3.97 (dt, *J* = 6.7, 9.7 Hz, 1H, H-1′b), 4.77 (d, *J* = 1.6 Hz, 1H, H-1), 5.54 (s, 2H, CH_2Ar), 6.80–6.96, 7.29 (m, 4H, C_6H_4), 7.79 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 27.2 (C-2′), 54.9 (CH_2Ar), 55.9 (OCH_3), 63.0 (C-6), 67.6 (C-1′), 68.7 (C-4), 72.2 (C-2), 72.7 (C-3), 74.9 (C-5), 101.7 (C-1), 114.8, 115.1, 121.2 (3C, C_6H_4), 124.2 (C_2N_3H–C5), 131.3 (C_6H_4), 138.4 (C_6H_4–C1), 146.7 (C_2N_3H–C4), 161.7 (C_6H_4–C3); HR-MS Calcd for $C_{18}H_{25}N_3NaO_7$ [M+Na]$^+$: 418.1590, Found: 418.1591.

4.1.7.15. [1-(4′-Nitrophenyl)-1,2,3-triazol-4-yl]ethyl α-D-mannopyranoside (6i). Prepared from **17i** (61 mg, 0.11 mmol) according to general procedure D. Yield: 37 mg, 86%.

$[\alpha]_D$ +44.4 (*c* 1.00, MeOH); ^{1}H NMR (500 MHz, CD_3OD): δ 3.13 (t, *J* = 6.5 Hz, 2H, H-2′), 3.42 (m, 1H, H-5), 3.60 (t, *J* = 9.5 Hz, 1H, H-4), 3.64–3.73 (m, 2H, H-3, H-6a), 3.78–3.88 (m, 3H, H-2, H-6b, H-1′a), 4.09 (dt, *J* = 6.6, 9.8 Hz, 1H, H-1′b), 4.83 (d, *J* = 1.5 Hz, 1H, H-1), 8.18, 8.48 (m, 4H, C_6H_4), 8.55 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 27.3 (C-2′), 63.1 (C-6), 67.4 (C-1′), 68.7 (C-4), 72.2 (C-2), 72.8 (C-3), 75.0 (C-5), 101.8 (C-1), 122.0 (2C, C_6H_4), 122.5 (C_2N_3H–C5), 126.6 (2C, C_6H_4), 142.8 (C_6H_4–C1), 148.1 (C_2N_3H–C4), 148.9 (C_6H_4–C4); HR-MS Calcd for $C_{16}H_{20}N_4NaO_8$ [M+Na]$^+$: 419.1179, Found: 419.1177.

4.1.7.16. [1-(Pyridin-4′-yl)-1,2,3-triazol-4-yl]ethyl α-D-mannopyranoside (6j). Prepared from **17j** (63 mg, 0.12 mmol) according to general procedure D. Yield: 31 mg, 73%.

$[\alpha]_D$ +48.3 (*c* 1.00, MeOH); ^{1}H NMR (500 MHz, CD_3OD): δ 3.12 (t, *J* = 6.5 Hz, 2H, H-2′), 3.43 (m, 1H, H-5), 3.61 (t, *J* = 9.5 Hz, 1H, H-4),

3.64–3.73 (m, 2H, H-3, H-6a), 3.77–3.87 (m, 3H, H-2, H-6b, H-1′a), 4.08 (dt, *J* = 6.6, 9.8 Hz, 1H, H-1′b), 4.83 (d, *J* = 1.5 Hz, 1H, H-1), 7.99 (dd, *J* = 1.6, 4.8 Hz, 2H, C_5H_4N), 8.58 (s, 1H, C_2N_3H), 8.74 (m, 2H, C_5H_4N); ^{13}C NMR (125 MHz, CD_3OD): δ 27.2 (C-2′), 63.1 (C-6), 67.3 (C-1′), 68.7 (C-4), 72.2 (C-2), 72.7 (C-3), 75.0 (C-5), 101.8 (C-1), 115.5 (2C, C_5H_4N), 122.0 (C_2N_3H–C5), 145.3 (C_5H_4N–C1), 148.1 (C_2N_3H–C4), 152.4 (2C, C_5H_4N); HR-MS Calcd for $C_{15}H_{21}N_4O_6$ $[M+H]^+$: 353.1461, Found: 353.1460.

4.1.7.17. [1-(4′-Fluorophenyl)-1,2,3-triazol-4-yl]ethyl α-D-mannopyranoside (6k). Prepared from **17k** (65 mg, 0.12 mmol) according to general procedure D. Yield: 40 mg, 90%.

$[\alpha]_D$ +50.7 (*c* 1.00, MeOH); ^1H NMR (500 MHz, CD_3OD): δ 3.10 (t, *J* = 6.5 Hz, 2H, H-2′), 3.43 (m, 1H, H-5), 3.61 (t, *J* = 9.5 Hz, 1H, H-4), 3.65–3.74 (m, 2H, H-3, H-6a), 3.77–3.87 (m, 3H, H-2, H-6b, H-1′a), 4.07 (dt, *J* = 6.6, 9.8 Hz, 1H, H-1′b), 4.82 (d, *J* = 1.6 Hz, 1H, H-1), 7.34, 7.88 (m, 4H, C_6H_4), 8.33 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 27.2 (C-2′), 63.1 (C-6), 67.5 (C-1′), 68.7 (C-4), 72.2 (C-2), 72.8 (C-3), 75.0 (C-5), 101.8 (C-1), 117.8 (d, *J* = 23.8 Hz, 2C, C_6H_4), 122.6 (C_2N_3H–C5), 124.0 (d, *J* = 8.8 Hz, 2C, C_6H_4), 135.0 (d, *J* = 2.5 Hz, C_6H_4–C1), 147.4 (C_2N_3H–C4), 164.1 (d, *J* = 246.3 Hz, C_6H_4–C4); HR-MS Calcd for $C_{16}H_{20}FN_3NaO_6$ $[M+Na]^+$: 392.1234, Found: 392.1238.

4.1.7.18. 1-(α-D-Mannopyranosyl)-4-phenyl-1,2,3-triazole (7n). Prepared from **20n** (50 mg, 0.11 mmol) according to general procedure D. Yield: 29 mg, 89%.

$[\alpha]_D$ +98.0 (*c* 1.34, MeOH); ^1H NMR (500 MHz, CD_3OD): δ 3.38 (ddd, *J* = 2.5, 6.6, 8.9 Hz, 1H, H-5), 3.76–3.80 (m, 2H, H-4, H-6a), 3.85 (dd, *J* = 2.5, 12.1 Hz, 1H, H-6b), 4.12 (dd, *J* = 3.5, 8.5 Hz, 1H, H-3), 4.76 (t, *J* = 3.1 Hz, 1H, H-2), 6.08 (d, *J* = 2.7 Hz, 1H, H-1), 7.34–7.38, 7.43–7.46, 7.84–7.85 (m, 5H, C_6H_5), 8.51 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 62.6 (C-6), 68.7 (C-4), 70.1 (C-2), 72.6 (C-3), 78.7 (C-5), 88.5 (C-1), 122.1 (C_2N_3H–C5), 126.8, 129.5, 130.0, 131.4 (6C, C_6H_5), 149.0 (C_2N_3H–C4); HR-MS Calcd for $C_{14}H_{17}N_3NaO_5$ $[M+Na]^+$: 330.1066, Found: 330.1060.

4.1.7.19. 1-(α-D-Mannopyranosyl)-4-(4-methylphenyl)-1,2,3-triazole (7o). Prepared from **20o** (46 mg, 94 μmol) according to general procedure D. Yield: 20 mg, 65%.

$[\alpha]_D$ +84.6 (*c* 0.63, MeOH); ^1H NMR (500 MHz, CD_3OD): δ 2.36 (s, 3H, PhC*H*$_3$), 3.36 (ddd, *J* = 2.3, 6.7, 8.9 Hz, 1H, H-5), 3.74–3.78 (m, 2H, H-4, H-6a), 3.84 (dd, *J* = 2.4, 12.1 Hz, 1H, H-6b), 4.10 (dd, *J* = 3.5, 8.5 Hz, 1H, H-3), 4.74 (t, *J* = 3.0 Hz, 1H, H-2), 6.05 (d, *J* = 2.6 Hz, 1H, H-1), 7.25, 7.72 (AA′, BB′ of AA′BB′, *J* = 8.0 Hz, 4H, C_6H_4), 8.45 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 21.3 (PhCH_3), 62.6 (C-6), 68.7 (C-4), 70.1 (C-2), 72.6 (C-3), 78.7 (C-5), 88.5 (C-1), 121.7 (C_2N_3H–C5), 126.7, 128.6, 130.6, 139.6 (6C, C_6H_5), 149.2 (C_2N_3H–C4); HR-MS Calcd for $C_{15}H_{18}N_3NaO_5$ $[M+Na]^+$: 344.1222, Found: 344.1215.

4.1.7.20. 4-(3-Chlorophenyl)-1-(α-D-mannopyranosyl)-1,2,3-triazole (7p). Prepared from **20p** (43 mg, 84 μmol) according to general procedure D. Yield: 25 mg, 87%.

$[\alpha]_D$ +89.2 (*c* 0.50, MeOH); ^1H NMR (500 MHz, CD_3OD): δ 3.37 (ddd, *J* = 2.4, 6.7, 8.7 Hz, 1H, H-5), 3.77 (dd, *J* = 6.6, 12.2 Hz, 1H, H-6a), 3.78 (t, *J* = 8.6 Hz, 1H, H-4), 3.87 (dd, *J* = 2.4, 12.1 Hz, 1H, H-6b), 4.11 (dd, *J* = 3.5, 8.3 Hz, 1H, H-3), 4.75 (t, *J* = 3.1 Hz, 1H, H-2), 6.07 (d, *J* = 2.6 Hz, 1H, H-1), 7.37 (d, *J* = 8.1 Hz, 1H, C_6H_4-H6), 7.44 (t, *J* = 7.9 Hz, 1H, C_6H_4–H5), 7.79 (d, *J* = 7.7 Hz, 1H, C_6H_4–H4), 7.90 (s, 1H, C_6H_4-H2), 8.58 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 62.6 (C-6), 68.7 (C-4), 70.1 (C-2), 72.6 (C-3), 78.7 (C-5), 88.6 (C-1), 122.7 (C_2N_3H–C5), 125.0, 126.6, 129.4, 131.6, 133.5, 136.0 (C_6H_4), 147.7 (C_2N_3H–C4); HR-MS Calcd for $C_{14}H_{16}ClN_3NaO_5$ $[M+Na]^+$: 364.0676, Found: 364.0676.

4.1.7.21. 4-(4-Trifluoromethylphenyl)-1-(α-D-mannopyranosyl)-1,2,3-triazole (7q). Prepared from **20q** (46 mg, 85 μmol) according to general procedure D. Yield: 27 mg, 86%.

$[\alpha]_D$ +83.4 (*c* 0.34, MeOH); ^1H NMR (500 MHz, CD_3OD): δ 3.38 (m, 1H, H-5), 3.77–3.81 (m, 2H, H-4, H-6a), 3.85 (dd, *J* = 1.9, 12.0 Hz, 1H, H-6b), 4.11 (dd, *J* = 3.3, 8.4 Hz, 1H, H-3), 4.76 (t, *J* = 2.7 Hz, 1H, H-2), 6.10 (d, *J* = 2.3 Hz, 1H, H-1), 7.76, 8.06 (AA′, BB′ of AA′BB′, *J* = 8.0 Hz, 4H, C_6H_4), 8.66 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 62.6 (C-6), 68.7 (C-4), 70.1 (C-2), 72.6 (C-3), 78.8 (C-5), 88.6 (C-1), 123.2 (C_2N_3H–C5), 125.6 (q, *J* = 272 Hz, CF_3), 127.0 (q, *J* = 3.7 Hz, 2C, C_6H_4–C3, C5), 127.2 (2C, C_6H_4–C2, C6), 131.2 (d, *J* = 32.4 Hz, C_6H_4–C4), 135.5 (C_6H_4–C1), 147.6 (C_2N_3H–C4); HR-MS Calcd for $C_{15}H_{16}F_3N_3NaO_5$ $[M+Na]^+$: 398.0940, Found: 398.0942.

4.1.7.22. 1-(α-D-Mannopyranosyl)-4-(3-pyridyl)-1,2,3-triazole (7r). Prepared from **20r** (47 mg, 98 μmol) according to general procedure D. Yield: 28 mg, 92%.

$[\alpha]_D$ +86.7 (*c* 0.93, MeOH); ^1H NMR (500 MHz, CD_3OD): δ 3.40 (ddd, *J* = 2.3, 6.6, 8.6 Hz, 1H, H-5), 3.77–3.82 (m, 2H, H-4, H-6a), 3.86 (dd, *J* = 2.4, 12.1 Hz, 1H, H-6b), 4.11 (dd, *J* = 3.5, 8.4 Hz, 1H, H-3), 4.76 (t, *J* = 3.1 Hz, 1H, H-2), 6.11 (d, *J* = 2.7 Hz, 1H, H-1), 7.54 (dd, *J* = 5.0, 7.8 Hz, 1H, C_5H_4N-H5), 8.31 (m, 1H, C_5H_4N-H6), 8.53 (dd, *J* = 1.4, 4.9 Hz, 1H, C_5H_4N-H4), 8.68 (s, 1H, C_2N_3H), 9.04 (d, *J* = 1.5, 1H, C_5H_4N–H2); ^{13}C NMR (125 MHz, CD_3OD): δ 62.6 (C-6), 68.7 (C-4), 70.1 (C-2), 72.6 (C-3), 78.8 (C-5), 88.6 (C-1), 123.1 (C_2N_3H–C5), 125.6 (C_5H_4N–C5), 128.5 (C_5H_4N–C1), 135.0 (C_5H_4N–C6), 145.6 (C_2N_3H–C4), 147.3 (C_5H_4N–C2), 149.7 (C_5H_4N–C4); HR-MS Calcd for $C_{13}H_{16}N_4NaO_5$ $[M+Na]^+$: 331.1018, Found: 331.1013.

4.1.7.23. 1-(α-D-Mannopyranosyl)-4-phenoxymethyl-1,2,3-triazole (7s). Prepared from **20s** (46 mg, 90 μmol) according to general procedure D. Yield: 27 mg, 89%.

$[\alpha]_D$ +57.0 (*c* 0.90, MeOH); ^1H NMR (500 MHz, CD_3OD): δ 3.30 (m, 1H, H-5), 3.70–3.74 (m, 2H, H-4, H-6a), 3.78 (dd, *J* = 1.8, 12.1 Hz, 1H, H-6b), 4.04 (dd, *J* = 3.2, 8.4 Hz, 1H, H-3), 4.67 (m, 1H, H-2), 5.14 (s, 2H, C*H*$_2$OPh), 6.00 (d, *J* = 1.7 Hz, 1H, H-1), 6.91 (t, *J* = 7.3 Hz, 1H, C_6H_5-H4), 6.96 (d, *J* = 8.1 Hz, 2H, C_6H_5-H2, H6), 7.24 (t, *J* = 7.8 Hz, 2H, C_6H_5-H3, H5), 8.20 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 62.2 (*C*H$_2$OPh), 62.5 (C-6), 68.6 (C-4), 70.1 (C-2), 72.5 (C-3), 78.6 (C-5), 88.4 (C-1), 115.8 (2C, C_6H_5–C2, C6), 122.3 (C_6H_5–C4), 125.3 (C_2N_3H–C5), 130.5 (2C, C_6H_5–C3, C5), 145.4 (C_2N_3H–C4), 159.7 (C_6H_5–C1); HR-MS Calcd for $C_{15}H_{19}N_3NaO_5$ $[M+Na]^+$: 360.1172, Found: 360.1171.

4.1.7.24. 1-(α-D-Mannopyranosyl)methyl-4-phenyl-1,2,3-triazole (8n). Prepared from **25n** (38 mg, 78 μmol) according to general procedure D. Yield: 22 mg, 87%.

$[\alpha]_D$ +30.6 (*c* 0.91, MeOH); ^1H NMR (500 MHz, CD_3OD): δ 3.73–3.75 (m, 2H, H-4, H-6a), 3.79–3.85 (m, 3H, H-2, H-3, H-5), 3.88 (dd, *J* = 7.2, 11.5 Hz, 1H, H-6b), 4.25 (dt, *J* = 4.8, 7.9 Hz, 1H, H-1), 4.73 (dd, *J* = 8.0, 14.4 Hz, 1H, H-1′a), 4.76 (dd, *J* = 4.5, 14.3 Hz, 1H, H-1′b), 7.33 (t, *J* = 7.7 Hz, 1H, C_6H_5-H4), 7.42 (t, *J* = 7.8 Hz, 2H, C_6H_5-H3, H5), 7.81 (d, *J* = 8.0 Hz, 2H, C_6H_5-H2, H6), 8.45 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 50.9 (C-1′), 62.1 (C-6), 69.1 (C-2), 70.0 (C-4), 72.5 (C-3), 74.9 (C-1), 78.5 (C-5), 123.4 (C_2N_3H–C5), 126.7, 129.3, 129.9, 131.8 (6C, C_6H_5), 148.8 (C_2N_3H–C4); HR-MS Calcd for $C_{15}H_{19}NaN_3O_5$ $[M+Na]^+$: 344.1222, Found: 344.1222.

4.1.7.25. 1-(α-D-Mannopyranosyl)methyl-4-(4-methylphenyl)-1,2,3-triazole (8o). Prepared from **25o** (42 mg, 84 μmol) according to general procedure D. Yield: 25 mg, 87%.

$[\alpha]_D$ +33.8 (*c* 1.12, MeOH); ^1H NMR (500 MHz, CD_3OD): δ 2.35 (PhCH_3), 3.72–3.76 (m, 2H, H-4, H-6a), 3.80 (dt, *J* = 3.2, 7.1 Hz, 1H, H-5), 3.81–3.85 (m, 2H, H-2, H-3), 3.87 (dd, *J* = 7.1, 11.5 Hz,

1H, H-6b), 4.24 (dt, J = 5.1, 7.4 Hz, 1H, H-1), 4.72–4.75 (m, 2H, H-1′), 7.24, 7.69 (AA′, BB′ of AA′BB′, J = 8.0 Hz, 4H, C_6H_4), 8.40 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 21.3 (PhCH_3), 50.8 (C-1′), 62.0 (C-6), 69.1 (C-2), 69.9 (C-4), 72.4 (C-3), 75.0 (C-1), 78.4 (C-5), 123.0 (C_2N_3H–C5), 126.6, 128.9, 130.5, 139.3 (6C, C_6H_5), 148.9 (C_2N_3H–C4); HR-MS Calcd for $C_{16}H_{21}NaN_3O_5$ $[M+Na]^+$: 358.1379, Found: 358.1380.

4.1.7.26. 4-(3-Chlorophenyl)-1-(α-D-mannopyranosyl)methyl-1,2,3-triazole (8p). Prepared from **25p** (40 mg, 77 μmol) according to general procedure D. Yield: 23 mg, 83%.

$[\alpha]_D$ +31.5 (c 1.05, MeOH); ^{1}H NMR (500 MHz, CD_3OD): δ 3.73 (m, 1H, H-4), 3.74 (dd, J = 3.0, 11.5 Hz, 1H, H-6a), 3.79–3.82 (m, 2H, H-3, H-5), 3.83 (dd, J = 3.4, 8.7 Hz, 1H, H-2), 3.89 (dd, J = 7.4, 11.6 Hz, 1H, H-6b), 4.24 (dt, J = 4.7, 7.9 Hz, 1H, H-1), 4.73 (dd, J = 8.0, 14.4 Hz, 1H, H-1′a), 4.77 (dd, J = 4.4, 14.5 Hz, 1H, H-1′b), 7.33 (dd, J = 0.9, 8.1 Hz, 1H, C_6H_4-H6), 7.41 (t, J = 7.9 Hz, 1H, C_6H_4–H5), 7.74 (d, J = 7.8 Hz, 1H, C_6H_4–H4), 7.90 (t, J = 1.6 Hz, 1H, C_6H_4-H2), 8.51 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 50.9 (C-1′), 62.0 (C-6), 69.0 (C-2), 70.0 (C-4), 72.4 (C-3), 74.7 (C-1), 78.5 (C-5), 124.0 (C_2N_3H–C5), 124.9, 126.5, 129.1, 131.5, 133.9, 135.9 (6C, C_6H_4), 147.4 (C_2N_3H–C4); HR-MS Calcd for $C_{15}H_{18}ClNaN_3O_5$ $[M+Na]^+$: 378.0833, Found: 378.0833.

4.1.7.27. 4-(4-Trifluoromethylphenyl)-1-(α-D-mannopyranosyl)methyl-1,2,3-triazole (8q). Prepared from **25q** (47 mg, 84 μmol) according to general procedure D. Yield: 28 mg, 86%.

$[\alpha]_D$ +32.6 (c 1.03, MeOH); ^{1}H NMR (500 MHz, CD_3OD): δ 3.73–3.76 (m, 2H, H-4, H-6a), 3.81–3.86 (m, 3H, H-2, H-3, H-5), 3.89 (dd, J = 7.5, 11.5 Hz, 1H, H-6b), 4.25 (dt, J = 4.8, 7.9 Hz, 1H, H-1), 4.76 (dd, J = 8.0, 14.5 Hz, 1H, H-1′a), 4.79 (dd, J = 4.4, 14.5 Hz, 1H, H-1′b), 7.72, 8.01 (AA′, BB′ of AA′BB′, J = 8.2 Hz, 4H, C_6H_4), 8.60 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 51.0 (C-1′), 62.0 (C-6), 69.0 (C-2), 70.0 (C-4), 72.4 (C-3), 74.7 (C-1), 78.6 (C-5), 124.5 (C_2N_3H–C5), 125.6 (q, J = 271 Hz, CF_3), 126.9 (q, J = 3.7 Hz, 2C, C_6H_4–C3, C5), 127.0 (2C, C_6H_4–C2, C6), 130.9 (q, J = 32.4 Hz, C_6H_4–C4), 135.7 (C_6H_4–C1), 147.2 (C_2N_3H–C4); HR-MS Calcd for $C_{16}H_{18}F_3NaN_3O_5$ $[M+Na]^+$: 412.1096, Found: 412.1095.

4.1.7.28. 1-(α-D-Mannopyranosyl)methyl-4-(3-pyridyl)-1,2,3-triazole (8r). Prepared from **25r** (44 mg, 90 μmol) according to general procedure D. Yield: 24 mg, 83%.

$[\alpha]_D$ +31.2 (c 0.99, MeOH); ^{1}H NMR (500 MHz, CD_3OD): δ 3.71–3.74 (m, 2H, H-4, H-6a), 3.80–3.83 (m, 3H, H-2, H-3, H-5), 3.89 (dd, J = 7.7, 11.6 Hz, 1H, H-6b), 4.23 (dt, J = 4.6, 8.6 Hz, 1H, H-1), 4.76 (dd, J = 8.4, 14.4 Hz, 1H, H-1′a), 4.80 (dd, J = 4.2, 14.4 Hz, 1H, H-1′b), 7.53 (dd, J = 5.0, 7.9 Hz, 1H, C_5H_4N-H5), 8.28 (d, J = 8.0 Hz, 1H, C_5H_4N-H6), 8.51 (d, J = 4.8 Hz, 1H, C_5H_4N-H4), 8.63 (s, 1H, C_2N_3H), 9.02 (s, 1H, C_5H_4N–H2); ^{13}C NMR (125 MHz, CD_3OD): δ 48.1 (C-1′), 60.6 (C-6), 67.2 (C-2), 68.5 (C-4), 70.6 (C-3), 75.1 (C-1), 76.2 (C-5), 123.2 (C_2N_3H–C5), 124.5 (C_5H_4N–C5), 126.82 (C_5H_4N–C1), 134.3 (C_5H_4N–C6), 144.4 (C_2N_3H–C4), 145.7 (C_5H_4N–C2), 148.4 (C_5H_4N–C4); HR-MS Calcd for $C_{14}H_{18}NaN_4O_5$ $[M+Na]^+$: 345.1175, Found: 345.1175.

4.1.7.29. 1-(α-D-Mannopyranosyl)methyl-4-phenoxymethyl-1,2,3-triazole (8s). Prepared from **25s** (41 mg, 79 μmol) according to general procedure D. Yield: 23 mg, 83%.

$[\alpha]_D$ +22.8 (c 1.01, MeOH); ^{1}H NMR (500 MHz, CD_3OD): δ 3.69–3.76 (m, 3H, H-4, H-5, H-6a), 3.79–3.82 (m, 2H, H-2, H-3), 3.83 (dd, J = 6.5, 11.5 Hz, 1H, H-6b), 4.19 (dt, J = 5.0, 7.0 Hz, 1H, H-1), 4.69 (dd, J = 7.5, 14.5 Hz, 1H, H-1′a), 4.72 (dd, J = 5.0, 14.5 Hz, 1H, H-1′b), 5.15 (s, 2H, CH_2OPh), 6.94 (t, J = 7.4 Hz, 1H, C_6H_5-H4), 7.00 (d, J = 8.1 Hz, 2H, C_6H_5-H2, H6), 7.27 (m, 2H, C_6H_5-H3, H5), 8.17 (s, 1H, C_2N_3H); ^{13}C NMR (125 MHz, CD_3OD): δ 50.9 (C-1′), 62.0 (C-6), 62.3 (CH_2OPh), 69.0 (C-2), 69.8 (C-4), 72.4 (C-3), 74.9 (C-1), 78.4 (C-5), 115.9 (2C, C_6H_5–C2, C6), 122.2 (C_6H_5–C4), 126.4 (C_2N_3H–C5), 130.5 (2C, C_6H_5–C3, C5), 145.0 (C_2N_3H–C4), 159.8 (C_6H_5–C1); HR-MS Calcd for $C_{16}H_{21}N_3NaO_5$ $[M+Na]^+$: 374.1328, Found: 374.1328.

4.2. Biological evaluation

4.2.1. Competitive binding assay

A recombinant protein consisting of the CRD of FimH linked with a thrombin cleavage site to a 6His-tag (FimH-CRD-Th-6His) was expressed in *E. coli* strain HM125 and purified by affinity chromatography.[45] To determine the affinity of the various FimH antagonists, an competitive binding assay described previously[45] was applied. Microtiter plates (F96 MaxiSorp, Nunc) were coated with 100 μL/well of a 10 μg/mL solution of FimH-CRD-Th-6His in 20 mM HEPES, 150 mM NaCl and 1 mM $CaCl_2$, pH 7.4 (assay buffer) overnight at 4 °C. The coating solution was discarded and the wells were blocked with 150 μL/well of 3% BSA in assay buffer for 2 h at 4 °C. After three washing steps with assay buffer (150 μL/well), a four-fold serial dilution of the test compound (50 μL/well) in assay buffer containing 5% DMSO and streptavidin-peroxidase coupled TM-PAA polymer (50 μL/well of a 0.5 μg/mL solution) were added. On each individual microtiter plate *n*-heptyl α-D-mannopyranoside (**1b**) was tested in parallel. The plates were incubated for 3 h at 25 °C and 350 rpm and then carefully washed four times with 150 μL/well assay buffer. After the addition of 100 μL/well of ABTS-substrate, the colorimetric reaction was allowed to develop for 4 min, then stopped by the addition of 2% aqueous oxalic acid before the optical density (OD) was measured at 415 nm on a microplate-reader (Spectramax 190, Molecular Devices, California, USA). The IC_{50} values of the compounds tested in duplicates were calculated with prism software (GraphPad Software, Inc., La Jolla, USA). The IC_{50} defines the molar concentration of the test compound that reduces the maximal specific binding of TM-PAA polymer to FimH-CRD by 50%. The relative IC_{50} (rIC_{50}) is the ratio of the IC_{50} of the test compound to the IC_{50} of **1b**.

4.2.2. Aggregometry assay

The aggregometry assay was carried out as previously described.[46] In short, the percentage of aggregation of *E. coli* UTI89[59] (UTI89wt) with guinea pig erythrocytes (GPE) was quantitatively determined by measuring the optical density at 740 nm and 37 °C under stirring at 1000 rpm using an APACT 4004 aggregometer (Endotell AG, Allschwil, Switzerland). GPE were separated from guinea pig blood (Charles River Laboratories, Sulzfeld, Germany) using Histopaque (density of 1.077 g/mL at 24 °C, Sigma-Aldrich, Buchs, Switzerland). Prior to the measurements, the cell densities of *E. coli* and GPE were adjusted to an OD_{600} of 4, corresponding to 1.9×10^8 CFU/mL and 2.2×10^6 cells/mL respectively. For the calibration of the instrument, the aggregation of protein poor plasma (PPP) using PBS alone was set as 100% and the aggregation of protein rich plasma (PRP) using GPE as 0%. After calibration, measurements were performed with 250 μL GPE and 50 μL bacterial suspension and the aggregation monitored over 600 s. After the aggregation phase of 600 s, 25 μL of antagonist in PBS were added to each cuvette and disaggregation was monitored for 1400 s. UTI89 Δ*fimA-H* was used as negative control.

4.3. Determination of the pharmacokinetic parameters

4.3.1. Materials

Dimethyl sulfoxide (DMSO) and 1-octanol were purchased from Sigma-Aldrich (St. Louis MI, USA). PAMPA System Solution, GIT-0 Lipid Solution, and Acceptor Sink Buffer were ordered from pIon (Woburn MA, USA). Acetonitrile (MeCN) was bought from Acros Organics (Geel, Belgium).

4.3.2. LC–MS measurements

Analyses were performed using an Agilent 1100/1200 Series HPLC System coupled to a 6410 Triple Quadrupole mass detector (Agilent Technologies, Inc., Santa Clara CA, USA) equipped with electrospray ionization. The system was controlled with the Agilent MassHunter Workstation Data Acquisition software (version B.01.04). The column used was an Atlantis® T3 C18 column (2.1 x 50 m) with a 3 μm particle size (Waters Corp., Milford MA, USA). The mobile phase consisted of two eluents: solvent A (H_2O, containing 0.1% formic acid, v/v) and solvent B (MeCN, containing 0.1% formic acid, v/v), both delivered at 0.6 mL/min. The gradient was ramped from 95% A/5% B to 5% A/95% B over 1 min, and then hold at 5% A/95% B for 0.1 min. The system was then brought back to 95% A/5% B, resulting in a total duration of 4 min. MS parameters such as fragmentor voltage, collision energy and polarity were optimized individually for each compound, and the molecular ion was followed for each compound in the multiple reaction monitoring mode. The concentrations of the analytes were quantified by the Agilent Mass Hunter Quantitative Analysis software (version B.01.04).

4.3.3. log $D_{7.4}$ determination

The *in silico* prediction tool ALOGPS[60] was used to estimate the log *P* values. Depending on these values, the compounds were classified into three categories: hydrophilic compounds (log *P* below zero), moderately lipophilic compounds (log *P* between zero and one) and lipophilic compounds (log *P* above one). For each category, two different ratios (volume of 1-octanol to volume of buffer) were defined as experimental parameters:

Compound type	log *P*	Ratios (1-octanol:buffer)
Hydrophilic	<0	30:140, 40:130
Moderately lipophilic	0–1	70:110, 110:70
Lipophilic	>1	3:180, 4:180

Equal amounts of phosphate buffer (0.1 M, pH 7.4) and 1-octanol were mixed and shaken vigorously for 5 min to saturate the phases. The mixture was left until separation of the two phases occurred, and the buffer was retrieved. Stock solutions of the test compounds were diluted with buffer to a concentration of 1 μM. For each compound, six determinations, that is, three determinations per 1-octanol : buffer ratio, were performed in different wells of a 96-well plate. The respective volumes of buffer containing analyte (1 μM) were pipetted to the wells and covered by saturated 1-octanol according to the chosen volume ratio. The plate was sealed with aluminium foil, shaken (1350 rpm, 25 °C, 2 h) on a Heidoph Titramax 1000 plate-shaker (Heidolph Instruments GmbH & Co. KG, Schwabach, Germany) and centrifuged (2000 rpm, 25 °C, 5 min, 5804 R Eppendorf centrifuge, Hamburg, Germany). The aqueous phase was transferred to a 96-well plate for analysis by liquid chromatography-mass spectrometry (LC-MS).

log $D_{7.4}$ was calculated from the 1-octanol : buffer ratio (o:b), the initial concentration of the analyte in buffer (1 μM), and the concentration of the analyte in buffer (c_B) with equilibration:

$$\log D_{7.4} = \left(\frac{1\mu M - c_B}{c_B} \times \frac{1}{o:b} \right)$$

The average of the three log $D_{7.4}$ values per 1-octanol:buffer ratio was calculated. If the two mean values obtained for a compound did not differ by more than 0.1 unit, the results were accepted.

4.3.4. Parallel artificial membrane permeation assay (PAMPA)

log P_e was determined in a 96-well format with the PAMPA[50] permeation assay. For each compound, measurements were performed at three pH values (5.0, 6.2 and 7.4) in quadruplicates. For this purpose, 12 wells of a deep well plate, that is, four wells per pH-value, were filled with 650 μL PAMPA System Solution. Samples (150 μL) were withdrawn from each well to determine the blank spectra by UV-spectroscopy (SpectraMax 190, Molecular Devices, Silicon Valley Ca, USA). Then, analyte dissolved in DMSO was added to the remaining PAMPA System Solution to yield 50 μM solutions. To exclude precipitation, the optical density was measured at 650 nm, with 0.01 being the threshold value. Solutions exceeding this threshold were filtrated. Afterwards, samples (150 μL) were withdrawn to determine the reference spectra. Further 200 μL were transferred to each well of the donor plate of the PAMPA sandwich (pIon, Woburn MA, USA, P/N 110 163). The filter membranes at the bottom of the acceptor plate were impregnated with 5 μL of GIT-0 Lipid Solution and 200 μL of Acceptor Sink Buffer were filled into each acceptor well. The sandwich was assembled, placed in the GutBox™, and left undisturbed for 16 h. Then, it was disassembled and samples (150 μL) were transferred from each donor and acceptor well to UV-plates. Quantification was performed by both UV-spectroscopy and LC-MS. log P_e-values were calculated with the aid of the PAMPA Explorer Software (pIon, version 3.5).

4.3.5. Thermodynamic solubility

Microanalysis tubes (Labo-Tech J. Stofer LTS AG, Muttenz, Switzerland) were charged with 1 mg of solid substance and 250 μL of phosphate buffer (50 mM, pH 6.5). The samples were briefly shaken by hand, then sonicated for 15 min and vigorously shaken (600 rpm, 25 °C, 2 h) on a Eppendorf Thermomixer comfort. Afterwards, the samples were left undisturbed for 24 h. After measuring the pH, the saturated solutions were filtered through a filtration plate (MultiScreen® HTS, Millipore, Billerica MA, USA) by centrifugation (1500 rpm, 25 °C, 3 min). Prior to concentration determination by LC-MS, the filtrates were diluted (1:1, 1:10 and 1:100 or, if the results were outside of the calibration range, 1:1000 and 1:10000). The calibration was based on six values ranging from 0.1 to 10 μg/mL.

Acknowledgement

We thank the Swiss National Science Foundation (project K-32KI-120904) for their support.

Supplementary data

Supplementary data (HRMS and HPLC data of target compounds **5**–**8**) associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2011.08.057.

References and notes

1. Wiles, T. J.; Kulesus, R. R.; Mulvey, M. A. *Exp. Mol. Pathol.* **2008**, *85*, 11.
2. Fihn, S. D. *N. Engl. J. Med.* **2003**, *349*, 259.
3. Hooton, T. M. *Int. J. Antimicrob. Agents* **2001**, *17*, 259.
4. Gouin, S. G.; Wellens, A.; Bouckaert, J.; Kovensky, J. *ChemMedChem* **2009**, *4*, 749.
5. Rosen, D. A.; Hung, C. S.; Kline, K. A.; Hultgren, S. J. *Infect. Immun.* **2008**, *76*, 4290.
6. Capitani, G.; Eidam, O.; Glockshuber, R.; Grutter, M. G. *Microbes Infect.* **2006**, *8*, 2284.
7. Mulvey, M. A. *Cell Microbiol.* **2002**, *4*, 257.
8. Choudhury, D.; Thompson, A.; Stojanoff, V.; Langermann, S.; Pinkner, J.; Hultgren, S. J.; Knight, S. D. *Science* **1999**, *285*, 1061.
9. Bouckaert, J.; Berglund, J.; Schembri, M.; Genst, E. D.; Cools, L.; Wuhrer, M.; Hung, C. S.; Pinkner, J.; Slättergård, R.; Zavialov, A.; Choudhury, D.; Langermann, S.; Hultgren, S. J.; Wyns, L.; Klemm, P.; Oscarson, S.; Knight, S. D.; Greve, H. D. *Mol. Microbiol.* **2005**, *55*, 441.
10. Wellens, A.; Garofalo, C.; Nguyen, H.; Van Gerven, N.; Slättergård, R.; Hernalsteens, J.-P.; Wyns, L.; Oscarson, S.; De Greve, H.; Hultgren, S.; Bouckaert, J. *PLoS ONE* **2008**, *3*, 4.

11. Han, Z.; Pinker, J. S.; Ford, B.; Obermann, R.; Nolan, W.; Wildman, S. A.; Hobbs, D.; Ellenberger, T.; Cusumano, C. K.; Hultgren, S. J.; Janetka, J. W. *J. Med. Chem.* **2010**, *53*, 4779.
12. Langermann, S.; Mollby, R.; Burlein, J. E.; Palaszynski, S. R.; Auguste, C. G.; DeFusco, A.; Strouse, R.; Schenerman, M. A.; Hultgren, S. J.; Pinkner, J. S.; Winberg, J.; Guldevall, L.; Soderhall, M.; Ishikawa, K.; Normark, S.; Koenig, S. *J. Infect. Dis.* **2000**, *181*, 774.
13. Langermann, S.; Palaszynski, S.; Barnhart, M.; Auguste, G.; Pinkner, J. S.; Burlein, J.; Barren, P.; Koenig, S.; Leath, S.; Jones, C. H.; Hultgren, S. J. *Science* **1997**, *276*, 607.
14. Bouckaert, J.; Mackenzie, J.; de Paz, J. L.; Chipwaza, B.; Choudhury, D.; Zavialov, A.; Mannerstedt, K.; Anderson, J.; Pierard, D.; Wyns, L.; Seeberger, P. H.; Oscarson, S.; De Greve, H.; Knight, S. D. *Mol. Microbiol.* **2006**, *61*, 1556.
15. (a) Sharon, N. *Biochim. Biophys. Acta* **2006**, *1760*, 527; (b) Ofek, I.; Hasty, D. L.; Sharon, N. *FEMS Immun. Med. Microbiol.* **2003**, *38*, 181; (c) Sharon, N.; Ofek, I. In *Protein-Carbohydrate Interactions in Infectious Diseases*; Bewley, C. A., Ed.; RSC Biomolecular Sciences: Cambridge, UK, 2006; pp 49–72.
16. (a) Firon, N.; Ofek, I.; Sharon, N. *Biochem. Biophys. Res. Commun.* **1982**, *105*, 1426; (b) Firon, N.; Ofek, I.; Sharon, N. *Carbohydr. Res.* **1983**, *120*, 235; (c) Sharon, N. *FEBS Lett.* **1987**, *217*, 145.
17. (a) Neeser, J.-R.; Koellreutter, B.; Wuersch, P. *Infect. Immun.* **1986**, *52*, 428; (b) Lindhorst, T. K. *Top. Curr. Chem.* **2002**, *218*, 201. review; (c) Patel, A.; Lindhorst, T. K. *Carbohydr. Res.* **2006**, *341*, 1657; (d) Nagahori, N.; Lee, R. T.; Nishimura, S.-L.; Pagé, S.; Roy, R.; Lee, Y. C. *ChemBioChem* **2002**, *3*, 836; (e) Appeldoorn, C. C. M.; Joosten, J. A. F.; Maate, F. A.; Dobrindt, U.; Hacker, J.; Liskamp, R. M. J.; Khan, A. S.; Pieters, R. J. *Tetrahedron: Asymmetry* **2005**, *16*, 361; (f) Touaibia, M.; Wellens, A.; Shiao, T. C.; Wang, Q.; Sirois, S.; Bouckaert, J.; Roy, R. *ChemMedChem* **2007**, *2*, 1190.
18. (a) Firon, N.; Ashkenazi, S.; Mirelman, D.; Ofek, I.; Sharon, N. *Infect. Immun.* **1987**, *55*, 472; (b) Lindhorst, T. K.; Kötter, S.; Kubisch, J.; Krallmann-Wenzel, U.; Ehlers, S.; Kren, V. *Eur. J. Org. Chem.* **1998**, 1669; (c) Sperling, O.; Fuchs, A.; Lindhorst, T. K. *Org. Biomol. Chem.* **2006**, *4*, 3913; (d) Berglund, J.; Bouckaert, J.; De Greve, H.; Knight, S. PCT/US 2005/089733, 2005.
19. Klein, T.; Abgottspon, D.; Wittwer, M.; Rabbani, S.; Herold, J.; Jiang, X.; Kleeb, S.; Lüthi, C.; Scharenberg, M.; Bezençon, J.; Gubler, E.; Pang, L.; Smiesko, M.; Cutting, B.; Schwardt, O.; Ernst, B. *J. Med. Chem.* **2010**, *53*, 8627.
20. Hung, C. S.; Bouckaert, J.; Hung, D.; Pinkner, J.; Widberg, C.; Defusco, A.; Auguste, C. G.; Strouse, R.; Langermann, S.; Waksman, G.; Hultgren, S. J. *Mol. Microbiol.* **2002**, *44*, 903.
21. Ernst, B.; Magnani, J. L. *Nat. Rev. Drug Disc.* **2009**, *8*, 661.
22. Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
23. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.
24. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J., J.A.; Vreven, T.; Kudin, T. K. N.; Burant, J. C. et al. *Gaussian 2003*, Gaussian Inc: Pittsburgh, PA, 2003.
25. Sancho-Garcia, J. C.; Cornil, J. *J. Chem. Theory Comput.* **2005**, *1*, 581.
26. *Glide*, version 5.5; Schrödinger, LLC: New York, NY, 2009.
27. *Desmond Molecular Dynamics System*, version 2.2; D. E. Shaw Research: New York, NY, 2009.
28. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.
29. Kaufman, R. J.; Sidhu, R. S. *J. Org. Chem.* **1982**, *47*, 4941.
30. (a) Lamara, K.; Smalley, R. K. *Tetrahedron* **1991**, *47*, 2277; (b) Li, Y.; Gao, L.-X.; Han, F.-S. *Chem. Eur. J.* **2010**, *16*, 7969.
31. Hu, M.; Li, J.; Yao, S. Q. *Org. Lett.* **2008**, *10*, 5529.
32. (a) Sawanishi, H.; Tajima, K.; Tsuchiya, T. *Chem. Pharm. Bull.* **1987**, *35*, 4101; (b) Sapountzis, I.; Ettmayer, P.; Klein, C.; Mantoulidis, A.; Steegmaier, M.; Steurer, S.; Waizenegger, I. PCT/EP2008/058432, 2008.
33. (a) Tornøe, C. W.; Meldal, M. In *Peptides: The Wave of the Future: Proceedings of the Second International and the Seventeenth American Peptide Symposium*; Lebl, M., Houghten, R. A., Eds.; Springer, 2001; p 263; (b) Tornøe, C. W.; Christensen, C.; Meldal, M. *J. Org. Chem.* **2002**, *67*, 3057.
34. Kolb, H. C.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2001**, *40*, 2004.
35. Palomo, C.; Aizpurua, J. M.; Balentová, E.; Azcune, I.; Santos, J. I.; Jiménez-Barbero, J.; Cañada, J.; Miranda, J. I. *Org. Lett.* **2008**, *10*, 2227.
36. (a) Tietze, L.; Bothe, U. *Chem. Eur. J.* **1998**, *4*, 1179; (b) Gouin, S. G.; Vanquelef, E.; García Fernández, J. M.; Ortiz Mellet, C.; Dupradeau, F.-Y.; Kovensky, J. *J. Org. Chem.* **2007**, *72*, 9032.
37. Fernández-Megia, E.; Correa, J.; Rodríguez-Meizoso, I.; Riguera, R. *Macromolecules* **2006**, *39*, 2113.
38. Buckle, D. R.; Rockell, C. J. M. *J. Chem. Soc., Perkin Trans. I* **1982**, 627.
39. Stefely, J. A.; Palchaudhuri, R.; Miller, P. A.; Peterson, R. J.; Moraski, G. C.; Hergenrother, P. J.; Miller, M. J. *J. Med. Chem.* **2010**, *53*, 3389.
40. Benati, L.; Bencivenni, G.; Leardini, R.; Minozzi, M.; Nanni, D.; Scialpi, R.; Spagnolo, P.; Zanardi, G. *J. Org. Chem.* **2006**, *71*, 5822.
41. Ito, S.; Satoh, A.; Nagatomi, Y.; Hirata, Y.; Suzuki, G.; Kimura, T.; Satow, A.; Maehara, S.; Hikichi, H.; Hata, M.; Kawamoto, H.; Ohta, H. *Bioorg. Med. Chem.* **2008**, *16*, 9817.
42. (a) Kunz, H.; Pfrengle, W.; Rück, K.; Sager, W. *Synthesis* **1991**, *103*, 1039; (b) Györgydeak, Z.; Szilagyi, L.; Paulsen, H. *J. Carbohydr. Chem.* **1993**, *12*, 139.
43. (a) Myers, R. W.; Lee, Y. C. *Carbohydr. Res.* **1986**, *154*, 145; (b) Köll, P.; Förtsch, A. *Carbohydr. Res.* **1987**, *171*, 301.
44. Titz, A.; Radic, Z.; Schwardt, O.; Ernst, B. *Tetrahedron Lett.* **2006**, *47*, 2383.
45. Rabbani, S.; Jiang, X.; Schwardt, O.; Ernst, B. *Anal. Biochem.* **2010**, *407*, 188.
46. Abgottspon, D.; Rölli, G.; Hosch, L.; Steinhuber, A.; Jiang, X.; Schwardt, O.; Cutting, B.; Smiesko, M.; Jenal, U.; Ernst, B.; Trampuz, A. *J. Microbiol. Methods* **2010**, *82*, 249.
47. Giampapa, C. S.; Abraham, S. N.; Chiang, T. M.; Beachey, E. H. *J. Biol. Chem.* **1988**, *263*, 5362.
48. Aprikian, P.; Tchesnokova, V.; Kidd, B.; Yakovenko, O.; Yarov-Yarovoy, V.; Trinchina, E.; Vogel, V.; Thomas, W.; Sokurenko, E. *J. Biol. Chem.* **2007**, *282*, 23437.
49. Trong, I. L.; Aprikian, P.; Kidd, B. A.; Forero-Shelton, M.; Tchesnokova, V.; Rajagopal, P.; Rodriguez, V.; Interlandi, G.; Klevit, R.; Vogel, V.; Stenkamp, R. E.; Sokurenko, E. V.; Thomas, W. E. *Cell* **2010**, *141*, 645.
50. Kansy, M.; Senner, F.; Gubernator, K. *J. Med. Chem.* **1998**, *41*, 1007.
51. Dearden, J. C.; Bresnen, J. G. M. *QSAR Comb. Sci.* **1988**, *7*, 133.
52. Kerns, E. H. *J. Pharm. Sci.* **2001**, *90*, 1838.
53. Poulain, F.; Serre, A.-L.; Lalot, J.; Leclerc, E.; Quirion, J.-C. *J. Org. Chem.* **2008**, *73*, 2435.
54. Kessler, H.; Gehrke, M.; Griesinger, C. *Angew. Chem.* **1988**, *100*, 507; *Angew. Chem. Int. Ed.* **1988**, *27*, 490.
55. Reid, B. R.; Banks, K.; Flynn, P.; Nerdalt, W. *Biochemistry* **1989**, *28*, 10001.
56. Szilágyi, L.; Forgó, P. *Carbohydr. Res.* **1993**, *247*, 129.
57. Varma, M. V. S.; Feng, B.; Obach, R. S.; Troutman, M. D.; Chupka, J.; Miller, H. R.; El-Kattan, A. *J. Med. Chem.* **2009**, *52*, 4844.
58. Avdeef, A.; Bendels, S.; Di, L.; Faller, B.; Kansy, M.; Sugano, K.; Yamauchi, Y. *J. Pharm. Sci.* **2007**, *96*, 2893.
59. Mulvey, M. A.; Schilling, J. D.; Hultgren, S. J. *Infect. Immun.* **2001**, *69*, 4572.
60. VCCLAB, Virtual Computational Chemistry Laboratory, http://www.vcclab.org, 2005.; (b) Tetko, I. V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin, V. A.; Radchenko, E. V.; Zefirov, N. S.; Makarenko, A. S.; Tanchuk, V. Y.; Prokopenko, V. V. *J. Comput. Aid. Mol. Des.* **2005**, *19*, 453.

7.7 *ChemMedChem.* 2012; *7(8)*: 1404-22.

CHEMMEDCHEM

ChemPubSoc Europe

DOI: 10.1002/cmdc.201200125

VIP

FimH Antagonists: Structure–Activity and Structure–Property Relationships for Biphenyl α-D-Mannopyranosides

Lijuan Pang, Simon Kleeb, Katrin Lemme, Said Rabbani, Meike Scharenberg, Adam Zalewski, Florentina Schädler, Oliver Schwardt, and Beat Ernst*[a]

Urinary tract infections (UTIs) are caused primarily by uropathogenic *Escherichia coli* (UPEC), which encode filamentous surface-adhesive organelles called type 1 pili. FimH is located at the tips of these pili. The initial attachment of UPEC to host cells is mediated by the interaction of the carbohydrate recognition domain (CRD) of FimH with oligomannosides on urothelial cells. Blocking these lectins with carbohydrates or analogues thereof prevents bacterial adhesion to host cells and therefore offers a potential therapeutic approach for prevention and/or treatment of UTIs. Although numerous FimH antagonists have been developed so far, few of them meet the requirement for clinical application due to poor pharmacokinetics. Additionally, the binding mode of an antagonist to the CRD of FimH can switch from an in-docking mode to an out-docking mode, depending on the structure of the antagonist. In this communication, biphenyl α-D-mannosides were modified to improve their binding affinity, to explore their binding mode, and to optimize their pharmacokinetic properties. The inhibitory potential of the FimH antagonists was measured in a cell-free competitive binding assay, a cell-based flow cytometry assay, and by isothermal titration calorimetry. Furthermore, pharmacokinetic properties such as log*D*, solubility, and membrane permeation were analyzed. As a result, a structure–activity and structure–property relationships were established for a series of biphenyl α-D-mannosides.

Introduction

Urinary tract infections (UTIs), the most prevalent series of infectious diseases worldwide, affect millions of people and account for significant morbidity as well as high medical costs.[1] The primary cause of UTIs are strains of uropathogenic *Escherichia coli* (UPEC), which make up 70–95% of reported cases.[1a,2] UTIs are treated with antibiotics; however, recurrent infections by UPEC with subsequent antibiotic exposure can lead to the emergence of antimicrobial resistance.[3]

Adhesion to host cells is the initial step of microbial infection. To gain an initial foothold within the bladder, UPEC strains encode filamentous surface-adhesive organelles called type 1 pili (fimbriae).[4] They mediate bacterial attachment to uroplakin Ia, a glycoprotein located on urothelial cells. This initial step prevents the clearance of *E. coli* by the bulk flow of urine and facilitates the invasion of host cells.[1b,5] A bacterial lectin known as FimH is located at the tips of type 1 pili. The carbohydrate recognition domain (CRD) of this lectin is responsible for binding to the complementary carbohydrate epitope of the host tissue. Blocking this lectin by a carbohydrate or a glycomimetic thereof offers a potential therapeutic approach for prevention and/or treatment of UTIs.[6]

More than two decades ago, Sharon and co-workers explored various mannosides and oligomannosides as potential antagonists for type 1 pili-mediated bacterial adhesion and observed interactions in the micro- to millimolar range.[7] The first crystal structure of FimH was solved in 1999,[8] and since then, numerous crystallographic studies have been reported, greatly facilitating the design of high-affinity ligands.[9] In summary, the reported affinities can be rationalized on the basis of the structure of FimH: First, the binding pocket accommodates the mannose with the hydroxy groups forming an extended hydrogen bond network. Second, the entrance to the binding site, referred to as the "tyrosine gate", is formed by three hydrophobic amino acids (Tyr48, Tyr137, and Ile52)[9a] and can host aliphatic and aromatic aglycones.

As a consequence of hydrophobic contacts of the alkyl aglycone, *n*-heptyl α-D-mannopyranoside (**1**) exhibits nanomolar affinity.[9b] With aromatic aglycones such as **2**–**5** (Figure 1), further improvements in affinity were observed.[10] To explore the binding mode and to improve affinity as well as ADME properties, a series of biphenyl FimH antagonists were synthesized.

Results and Discussion

An unexpected docking mode was discovered upon co-crystallization of biphenyl mannoside **3** with the FimH CRD.[10d] Whereas the alkyl aglycone of *n*-butyl α-D-mannopyranoside[9b]

[a] *L. Pang,[+] S. Kleeb,[+] Dr. K. Lemme,[+] Dr. S. Rabbani,[+] Dr. M. Scharenberg, A. Zalewski, F. Schädler, Dr. O. Schwardt, Prof. Dr. B. Ernst*
Institute of Molecular Pharmacy, Pharmacenter, University of Basel
Klingelbergstrasse 50, 4056 Basel (Switzerland)
E-mail: beat.ernst@unibas.ch

[+] *These authors contributed equally to this work.*

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cmdc.201200125.

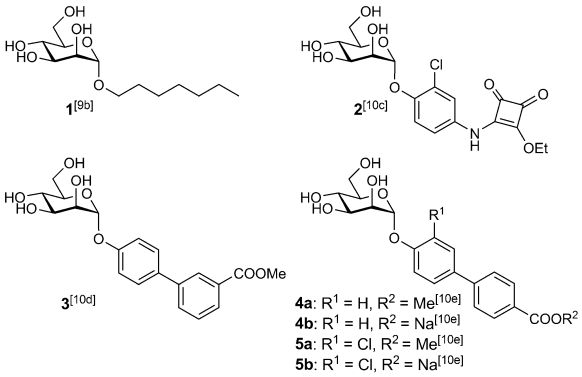


Figure 1. FimH antagonists: *n*-heptyl α-D-mannopyranoside (**1**) is used as reference compound; the squaric acid derivative **2** and biphenyl derivatives **3**–**5** exhibit nanomolar affinities.

interacts with both Tyr48 and Tyr137 of the tyrosine gate (in-docking mode),[10f] the biphenyl aglycone adopts the out-docking mode; that is, it interacts only with Tyr48 (Figure 2A), probably due to insufficient flexibility; π–π stacking of the outer aromatic ring of the biphenyl aglycone (ring B) with Tyr48 is effected by induced fit: a substantial move of Tyr48. Moreover, further stabilization of the protein–ligand complex by polar interaction between the ester in the *meta* position of **3** and the side chain of Arg98 was proposed.[10d]

In silico docking studies with biphenyl derivative **4a**[10e] suggested a similar out-docking mode (Figure 2B). A close inspection revealed empty space between the *ortho* position of the aromatic ring adjacent to the anomeric center (ring A) and the protein surface. Indeed, with an *ortho*-chloro substituent (→ **5a**, Figure 1), affinity was substantially improved. Further studies with FimH antagonists that exhibit enhanced flexibility (e.g., compound **6;** Figure 2C and Figure 3) indicated a switch from the out-docking mode to the in-docking mode. However, whether an optimal π–π stacking within the tyrosine gate can be realized remains to be determined. Finally, docking studies also indicated that elongation of the carboxylate-bearing *para* substituent enables a polar interaction between the carboxylate and Arg98 (e.g., compound **7**; Figure 2D and Figure 3).

Starting from antagonist **4**, we explored three types of modifications (Figure 3):

1) For optimizing the van der Waals contact between the *ortho* position of ring A and the binding pocket, a series of substituents — F, CH_3, CF_3, OCH_3, cyclopropyl, and CN — were introduced as depicted in Scheme 1.
2) To determine whether the out-docking mode reported for **3**[10d] results from insufficient flexibility, we increased the aglycone flexibility by introducing a methylene spacer between the anomeric oxygen and ring A of the biphenyl moiety (Scheme 2). This should decrease the conformational constraints to allow an optimized spatial arrangement of the aglycone in the tyrosine gate (→**6**, Figure 2C); at the same time, water solubility should be improved as a result of the decreased stacking tendency derived from disruption of the symmetry of the aglycone.[15]
3) To enable a polar interaction between the carboxylate substituent on ring B with Arg98 of FimH, we extended the *para* substituent of **4**, that is, we replaced it with either a flexible methyl ethanolate or a rigid methyl cyclopropanecarboxylate (Scheme 3). Biphenyl α-D-mannoside **24**[10d] shows a three- to eightfold lower affinity for FimH than its counterparts with a methoxycarbonyl substituent at the *meta* (→**3**)[10d] or *para* positions (→**4**)[10e] of ring B (Table 1). Han et al. assigned the increased affinity of compound **3** to a polar interaction of the ester with Arg98 of FimH.[10d] Because for spatial reasons the ester in the

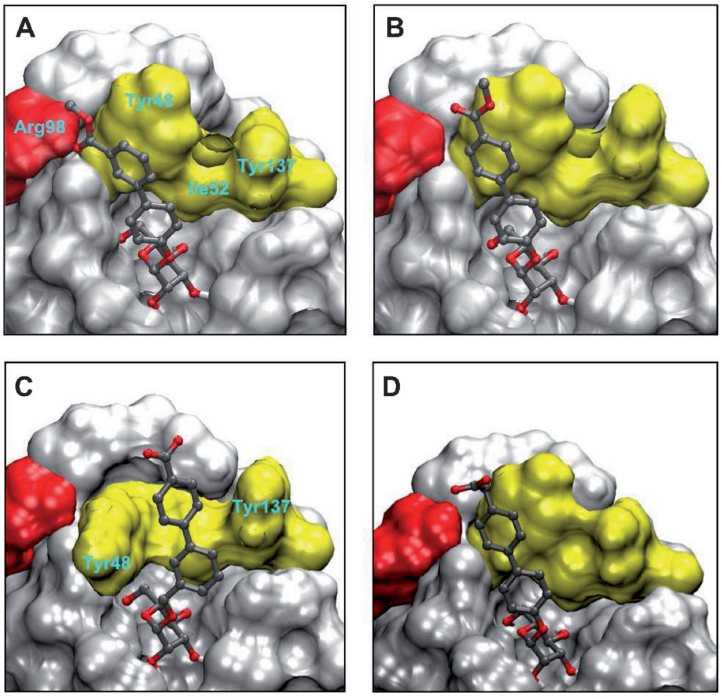


Figure 2. A) Crystal structure of biphenyl **3** (PDB ID: 3MCY)[10d] bound to the FimH CRD. B–D) In silico docking studies obtained with flexible docking (Glide software package)[11] to the same FimH CRD structure; top-scored binding mode of B) **4a**, C) **6**, and D) **7**.

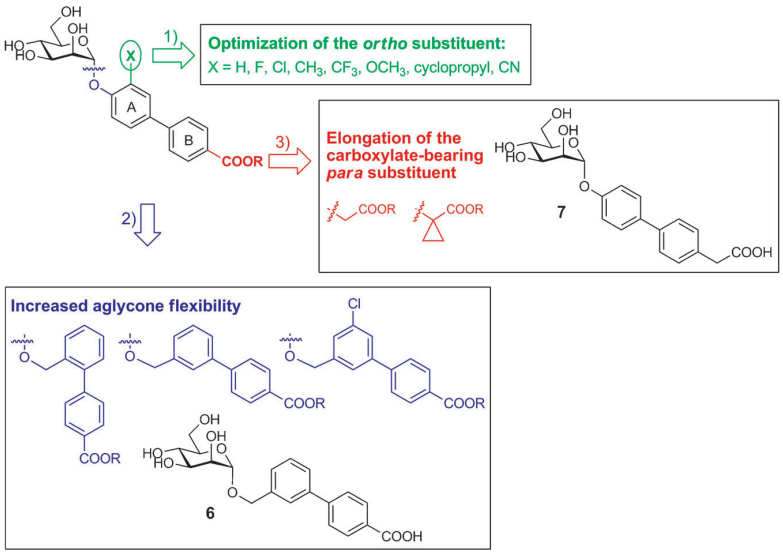


Figure 3. Modifications to the aglycone of FimH antagonists by 1) optimization of the *ortho* substituent, 2) an increase in the flexibility of the aglycone, and 3) elongation of the carboxylate-bearing *para* substituent.

para-substituted derivative **4** cannot establish a similar interaction with Arg98, the substantial improvement in affinity may result from solvation effects.

Synthesis

Optimization of ortho *substituents (Scheme 1)*

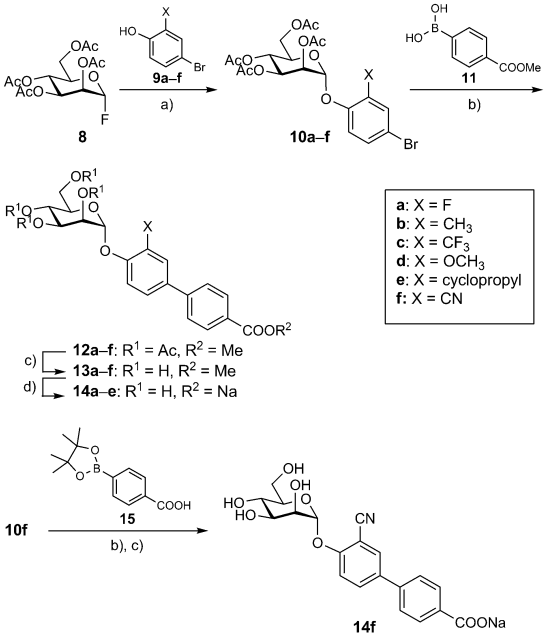
Mannosylation of phenols **9a–f** with mannosyl fluoride **8** and $BF_3{\cdot}OEt_2$ as promoter yielded α-mannosides **10a–f** stereospecifically.[12] Whereas the phenols **9a–d** and **9f** are commercially available, the cyclopropyl derivative **9e** was prepared via tandem carbolithiation/1,3-elimination according to Ocasio and Scanlan.[13] In a palladium-catalyzed Miyaura–Suzuki coupling[14] of **10a–f** with 4-methoxycarbonylphenylboronic acid (**11**), biphenyls **12a–f** were obtained in good to excellent yields. Deacetylation using Zemplén conditions (→**13a–f**) followed by saponification of the methyl esters gave the test compounds **14a–e**. Owing to the instability of the cyano group under aqueous basic conditions, **14f** was synthesized by coupling **10f** with 4-carboxyphenylboronic acid pinacol ester (**15**) followed by transesterification under Zemplén conditions to avoid the final saponification with aqueous sodium hydroxide.

Increase in aglycone flexibility (Scheme 2)

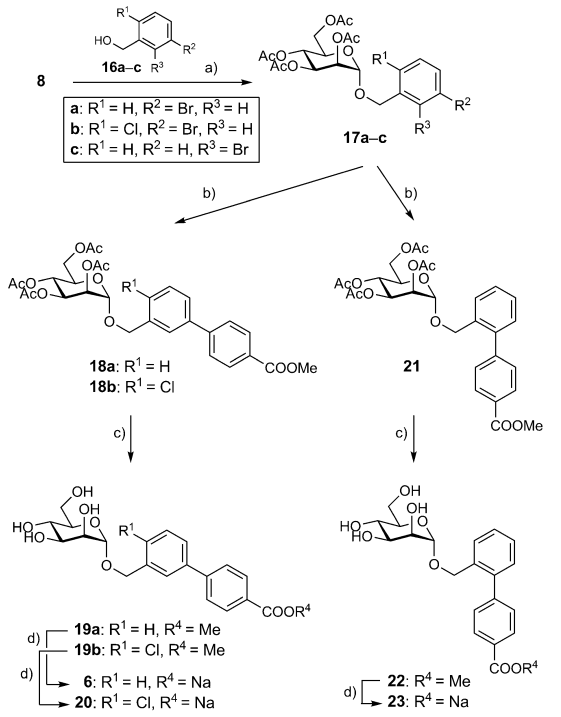
Benzyl alcohols **16a–c** were first mannosylated with donor **8**[12] to yield the benzyl mannosides **17a–c**. Subsequent cross-coupling with 4-methoxycarbonylphenylboronic acid (**11**) afforded acetates **18a,b** and **21**. Deacetylation of the mannose moiety (→**19a,b** and **22**) followed by saponification of the methyl esters gave compounds **6**, **20**, and **23**.

Elongation of the carboxylate-bearing para *substituent (Scheme 3)*

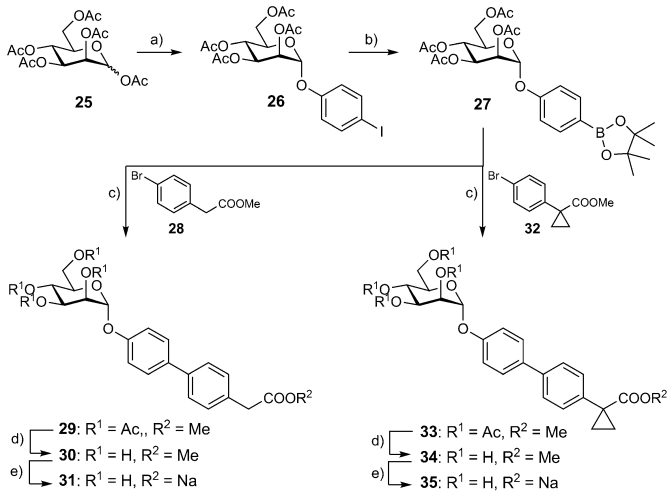
Peracetylated mannose **25** was treated with 4-iodophenol in the presence of $BF_3{\cdot}Et_2O$. The resulting iodide **26** was transformed into boronic acid pinacol ester **27**, which was coupled with 4-bromophenylacetic acid methyl ester (**28**) and 4-bromophenylcyclopropylcarboxylic acid methyl ester (**32**) under Miyaura–Suzuki coupling conditions[14] to yield biphenyls **29** and **33**. Deacetylation with sodium methoxide (→**30** and **34**) followed by saponification of the methyl ester yielded the sodium salts **31** and **35**.



Scheme 1. *Reagents and conditions:* a) $BF_3{\cdot}Et_2O$, CH_2Cl_2, 0 °C, 3 h (**10a–f**, 73–86%); b) $Pd(Cl_2)dppf{\cdot}CH_2Cl_2$, K_3PO_4, DMF, 80 °C, overnight (**12a–f**, 55–91%); c) NaOMe, MeOH, RT, 4 h (**13a–e**, **14f**, 52–73%); d) 1. 0.2 N $NaOH_{(aq)}$, MeOH, RT, overnight; 2. Dowex (Na^+), size-exclusion chromatography (P-2 gel) (**14a–e**, 15–74%).



Scheme 2. *Reagents and conditions:* a) $BF_3{\cdot}Et_2O$, CH_2Cl_2, 0 °C, 3 h (**17a–c**, 34–48%); b) 4-methoxycarbonylphenylboronic acid (**11**), $Pd(Cl_2)dppf{\cdot}CH_2Cl_2$, K_3PO_4, DMF, 80 °C, overnight (**18a,b** and **21**, 73–94%); c) NaOMe, MeOH, RT, 4 h (**19a,b** and **22**, 47–90%); d) 1. 0.2 N $NaOH_{(aq)}$, MeOH, RT, overnight; 2. Dowex (Na^+), size-exclusion chromatography (P-2 gel) (**6**, **20** and **23**, 10–96%).



Scheme 3. *Reagents and conditions:* a) 4-iodophenol, $BF_3{\cdot}Et_2O$, CH_2Cl_2, 40 °C, overnight (70%); b) bis(pinacolato)diborone, $Pd(Cl_2)dppf{\cdot}CH_2Cl_2$, KOAc, DMF, MW 120 °C, 2 h (50%); c) $Pd(Cl_2)dppf{\cdot}CH_2Cl_2$, K_3PO_4, DMF, 80 °C, overnight (34–56%); d) NaOMe, MeOH, RT, 4 h (33–95%); e) 1. 0.2 N $NaOH_{(aq)}$, MeOH, RT, overnight; 2. Dowex (Na^+), size-exclusion chromatography (P-2 gel) (**31**: 40%; **35**: 23%).

Binding affinity and activity

The biphenyl α-D-mannosides with varying *ortho* substituents (**5a–b**, **13a–f**, **14a–f**), increased aglycone flexibility (**6**, **19**, **20**, **22**, **23**), and elongated carboxylate-bearing *para* substituents (**30**, **31**, **34**, **35**) were evaluated in vitro by two competitive assay formats (Table 1). All antagonists were tested in a cell-free competitive binding assay.[16] Subsequently, the best candidates were investigated in a cell-based flow cytometry assay.[17]

The cell-free competitive binding assay is based on the interaction of a biotinylated polyacrylamide glycopolymer as competitor with the isolated CRD of FimH. In contrast, the cell-based flow cytometry assay involves the infection of human urinary bladder epithelial carcinoma cells with GFP-labeled UPECs expressing the complete type 1 pili (see the Experimental Section below for details). The competitors in the former assay are thus polymer-bound trimannosides, whereas in the latter the antagonists compete with more potent high-mannose oligosaccharides present on uroplakin Ia, which is located on the surface of human urinary bladder cells.[18,19] The interaction is further affected by the presence of high- and low-affinity states of the CRD of FimH. Aprikian et al. experimentally demonstrated that in full-length fimbriae, the pilin domain stabilizes the CRD domain in the low-affinity state, whereas the CRD domain alone adopts the high-affinity state.[20] Furthermore, it was recently shown that shear stress can induce a conformational switch (twist in the β-sandwich fold of the CRD domain), resulting in improved affinity.[21] Therefore, differing affinities were expected in the cell-based flow cytometry assay, in which full-length fimbriae are present, relative to the cell-free competitive binding assay.

Cell-free competitive binding assays[16]

These assays were performed twice for every compound with each concentration in duplicate. To ensure comparability between various antagonists, the reference compound *n*-heptyl α-D-mannopyranoside **1**[22] was tested each time in parallel. The affinities are reported relative to **1** as rIC_{50} in Table 1. A comparison of the affinities of compounds **4a** and **4b** with the *ortho*-substituted analogues **5a**, **13a–f** and **5b**, **14a–f** clearly demonstrates that *ortho* substituents on ring A indeed improve binding. However, the differences between the various substituted FimH antagonists are small. For a better understanding of these results, a more detailed analysis of the thermodynamic profile by isothermal titration calorimetry (ITC) was performed (see below). By increasing the flexibility of the aglycone, we expected a switch from the out-docking mode as present for antagonists **3** and **4** (Figure 2A,B) to the in-docking mode (represented by antagonist **6** in Figure 2C).[10f] However, affinities for all six representatives with increased spacer length between carbohydrate and aglycone (Table 1: **6**—, **19**, **20**, **22**, and **23**) were dramatically decreased. A similar tendency was observed

Table 1. Pharmacodynamic parameters of FimH antagonists.

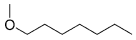
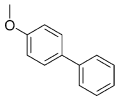
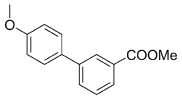
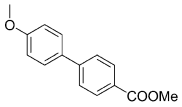
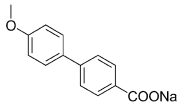
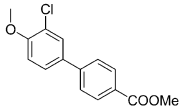
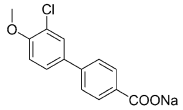
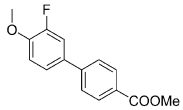
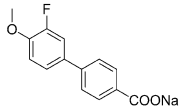
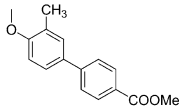
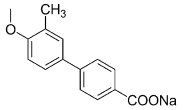
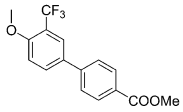
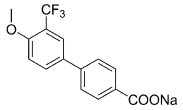
Compd	R	Binding assay IC_{50} [nM][a]	Binding assay rIC_{50}[b]	Flow cytometry IC_{50} [μM][a,c]
1[10e]		73 ± 7.9	1	3.9 ± 1.6
24[10d]		84.9	1.47	n.d.
3[10d]		28.6	0.55	n.d.
4a[10e]		10.4 ± 1.2	0.14	n.d.
4b[10e]		17.1 ± 2.2	0.15	n.d.
5a		4.8 ± 1.2	0.06	n.d.
5b		6.7 ± 2.1	0.09	0.33 ± 0.05
13a		8.0	0.14	n.d.
14a		33.5	0.58	1.54 ± 0.31
13b		23.3	0.40	n.d.
14b		9.2	0.16	1.83 ± 0.14
13c		2.6	0.04	n.d.
14c		8.9	0.15	0.89 ± 0.10

Table 1. (Continued)

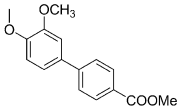
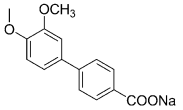
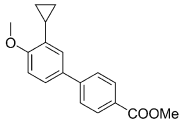
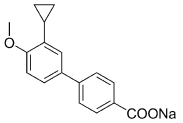
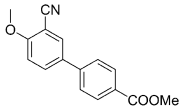
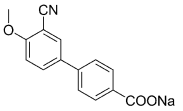
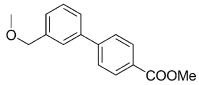
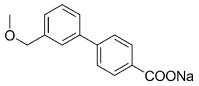
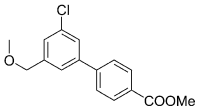
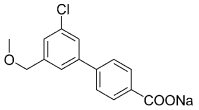
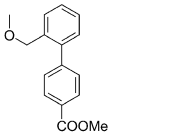
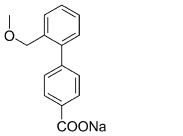
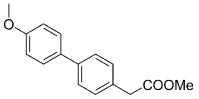
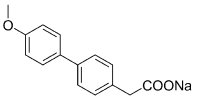
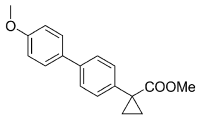
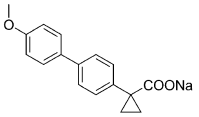
Compd	R	Binding assay IC_{50} [nM][a]	Binding assay rIC_{50}[b]	Flow cytometry IC_{50} [μM][a,c]
13d		3.5	0.06	n.d.
14d		4.8	0.08	1.95 ± 0.36
13e		31.7	0.55	n.d.
14e		63.0	1.09	4.85 ± 0.79
13f		22.5	0.39	n.d.
14f		33.9	0.58	n.d.
19a		56.1	0.97	n.d.
6		107.9	1.87	n.d.
19b		98.9	1.7	n.d.
20		142.2	2.44	n.d.
22		85.8	1.49	n.d.
23		642.0	11.14	n.d.

Table 1. (Continued)

Compd	R	Binding assay IC_{50} [nM][a]	Binding assay rIC_{50}[b]	Flow cytometry IC_{50} [μM][a,c]
30		63.2	1.09	n.d.
31		70.5	1.21	n.d.
34		49.5	0.85	n.d.
35		62.5	1.07	n.d.

[a] IC_{50} values were determined in a cell-free competitive binding assay.[16] [b] The rIC_{50} values were calculated by dividing the IC_{50} of the compound of interest by that of reference compound **1**; this leads to rIC_{50} values <1 for derivatives that bind better than reference **1**, and rIC_{50} values >1 for compounds with lower affinity than **1**. [c] The anti-adhesion potential to human epithelial bladder cells was determined in the flow cytometry assay;[17] n.d. = not determined.

for the biphenyls with an elongated carboxylate-bearing *para* substituent (Table 1: **30**, **31**, **34**, and **35**). It was previously described that the ester of **3** is placed within hydrogen bonding distance to form a polar interaction with Arg98 and Glu50.[10d] However, an improvement of affinity provided by a similar polar interaction between Arg98 and the antagonists **31** and **35** could not be achieved, probably due to the high desolvation penalty of Arg98. Finally, it is important to note that the free acids (sodium salt) of the antagonists in general showed slightly lower affinities than their methyl ester counterparts, with the only exceptions of **13b** and **14b** (Table 1). However, because the esters are thought to act as prodrugs and to be rapidly cleaved after oral application,[10e] the affinities of the carboxylates are relevant with regard to the therapeutic potential of the present FimH antagonists.

Cell-based flow cytometry assay[17]

These assays were performed in duplicate/triplicate, and *n*-heptyl α-D-mannopyranoside **1** was used as reference compound with an IC_{50} value of 3.9 ± 1.6 μM. The most potent antagonists **5b** and **14c** (Table 1) showed respective IC_{50} values of 0.33 ± 0.05 and 0.89 ± 0.10 μM. In general, the activities obtained from the flow cytometry assay were ~50-fold lower than the affinities determined in the target-based competitive assay (see above).

Isothermal titration calorimetry

Because the biological in vitro evaluation only revealed small differences between affinities, ITC experiments were carried out to study the thermodynamic profile of the variously *ortho*-substituted biphenyl compounds **5b** and **14a–f** in binding to FimH. ITC directly measures the heat of interaction (change in enthalpy, ΔH) at a constant temperature on titrating two compounds of known concentration that form an equilibrium complex.[23] It includes contributions from all equilibria that occur as the interacting molecules go from the free to the bound state, including those associated with solvent interactions and macromolecular conformational changes. The noncovalent interaction between a protein and a ligand can be quantified by the change in free energy (ΔG), consisting of the change in enthalpy (ΔH) and change in entropy (ΔS) [Eq. (1)].[24] The binding energy under standard conditions (ΔG°), in which all reactants and products are at a concentration of 1 M, can be calculated from the dissociation constant, K_D [Eq. (2)]. With ITC, K_D and ΔH can be measured directly, whereas ΔG and the entropy term $T\Delta S$ are calculated according to Equations (1) and (2).

$$\Delta G = \Delta H - T\Delta S \quad (1)$$

$$\Delta G = RT \ln K_D \quad (2)$$

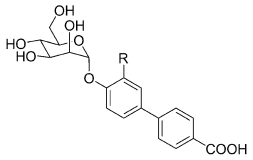
A favorable enthalpy term ΔH is associated with hydrogen bond formation, electrostatic, and dipole–dipole interactions at the overcompensation of the desolvation penalty.[25] The entropy term ΔS reflects the overall change in the degrees of freedom of a system. It can be dissected into translational and rigid body rotational entropy,[26] solvation entropy,[27] and entropy costs related to conformational changes of protein and ligand [Eq. (3)].[28] Whereas the formation of a protein–ligand complex is always associated with a decrease in translational and rotational freedom and therefore with entropy costs, the entropic contribution involving changes in solvation (ΔS_{solv}) and changes in rotational and vibrational entropy due to the loss of conformational flexibility (ΔS_{conf}) can differ both in sign and magnitude.[29]

$$\Delta S = \Delta S_{solv} + \Delta S_{trans/rot} + \Delta S_{conf} \quad (3)$$

The FimH CRD was used for the ITC experiments. It was prepared from FimH-CRD-Th-His_6 (see *Competitive binding assay*, Experimental Section below) by incubation with thrombin, as described earlier.[16]

The thermodynamic fingerprints of the various biphenyl derivatives (Table 2, Figure 4) reveal a significant improvement in the enthalpic term ($\Delta\Delta H$ −4.3 to −11.2 kJ mol^{-1}) for all substituted biphenyls (**5b**, **14a–f**) in comparison with the unsubstituted derivative **4b**. The largest enthalpy improvement was observed for the trifluoromethyl group (**14c**; Table 2). Interestingly, these largely improved enthalpic contributions are mostly compensated by entropic penalties ($-T\Delta\Delta S$ +3.2 to +7.5 kJ mol^{-1}), resulting in only marginally improved K_D values. In the best case, the trifluoromethyl derivative **14c**, a fourfold improvement in K_D was measured. Similar, but less pronounced

Table 2. Binding thermodynamics of FimH antagonists determined by ITC.



Compd	R	K_D [nM]	ΔG° [kJ mol^{-1}]	ΔH° [kJ mol^{-1}]	$-T\Delta S^\circ$ [kJ mol^{-1}]	N[a]	V_{vdW} [Å^3][b]
4b	H	14.1	−44.8	−47.3	+2.5	1.00	7.2
5b	Cl	3.7	−48.1	−55.5	+7.4	1.01	22.4
14a	F	9.2	−45.9	−51.6	+5.7	1.00	13.3
14b	Me	4.8	−47.5	−56.2	+8.7	1.01	26.7
14c	CF_3	3.2	−48.5	−58.5	+10.0	1.02	41.4
14d	OMe	7.7	−46.3	−52.5	+6.2	1.02	34.8
14e	*c*Pr	6.9	−46.6	−46.7	+0.1	1.01	52.5
14f	CN	7.4	−46.4	−55.0	+8.6	1.01	29.7

[a] Molar ratio of protein/ligand. [b] van der Waals volumes (V_{vdW}) of the *ortho* substituent were calculated with the Phase volCalc utility.[30]

effects were observed for most other *ortho* substituents. This trend was broken only by the cyclopropyl derivative **14e** ($\Delta\Delta H$ +0.7 kJ mol^{-1}, $-T\Delta\Delta S$ −2.4 kJ mol^{-1}; Table 2).

The influence of the *ortho* substituent on binding can be attributed to three factors. First, *ortho* substituents of appropriate volume establish an improved shape complementarity within the binding pocket, leading to a better van der Waals (vdW) contact and therefore an improvement in the enthalpy term ΔH. The improvement in enthalpy ($\Delta\Delta H$) correlates well with increasing vdW volumes of spherical *ortho* substituents (**5b**, **14a**–**c**; Figure 5). For non-spherical substituents (OMe, **14d**; cyclopropyl, **14e**; and CN, **14f**), the shape complementarity is not optimal, leading to only a decreased enthalpy contribution. However, better vdW contacts also lead to decreased conformational flexibility and therefore an entropic compensation by a less favorable ΔS_{conf} value. A second parameter is the desolvation enthalpy, which depends on the polarity of a specific *ortho* substituent and varies between −2.39 kJ mol^{-1} for CH_3 and 19.31 kJ mol^{-1} for CN.[25] Finally, depending on the surface area of the *ortho* substituent, the entropy of solvation may change. In summary, the various effects are superimposed and of opposing contributions to the free binding energy ΔG.

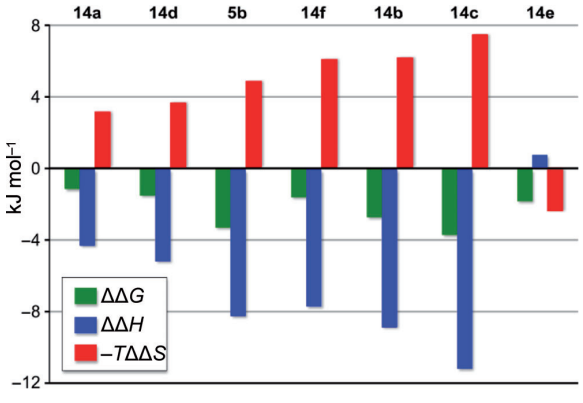


Figure 4. Enthalpy–entropy compensation, a property often reported for carbohydrate–lectin interactions,[31] for *ortho*-substituted biphenyl α-D-mannopyranosides; $\Delta\Delta G$, $\Delta\Delta H$, and $T\Delta\Delta S$ values for **5b** and **14a**–**f** are plotted relative to the unsubstituted derivative **4b**.

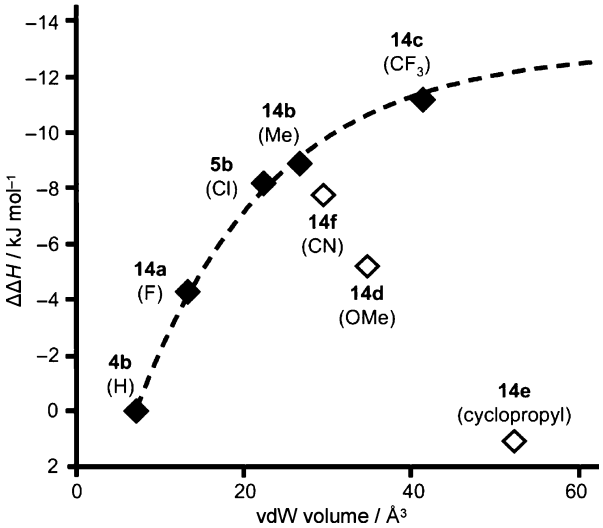


Figure 5. Correlation of $\Delta\Delta H$ (relative to antagonist **4b**) with the van der Waals volumes[30] of *ortho* substituents.

Physicochemical and in vitro pharmacokinetic characterization

To estimate the oral bioavailability and renal elimination of acids **4b**, **5b**, **6**, **14a**–**f**, **20**, **23**, **31**, **35**, and the methyl esters **4a**, **5a**, **13a**–**f**, **19a**–**b**, **22**, **30** and **34**, several physicochemical parameters (lipophilicity, solubility) as well as permeability through an artificial membrane and a cell monolayer were determined (Table 3). The free acids of the antagonists assessed in this study (**4b**, **5b**, **6**, **14a**–**f**, **20**, **23**, **31**, and **35**) are generally hydrophilic and soluble at pH values >5. All acids showed log $D_{7.4}$ values below zero and are therefore thought to undergo considerable renal clearance,[32] a prerequisite for FimH antagonists to reach their target in the urinary bladder. Permeation studies through an artificial membrane (PAMPA[33]) indicated for all acids except **14a** effective permeation values (log P_e) below −6.7, suggesting low absorption in the small intestine by passive permeation.[34] However, the high absorption potential of the fluoro-substituted biphenyl **14a** predicted by

Table 3. Physicochemical and in vitro pharmacokinetic parameters of FimH antagonists.

Compd	PAMPA log P_e [log 10^{-6} cm s^{-1}]/pH[a]	Caco-2 P_{app} [10^{-6} cm s^{-1}][b]			log $D_{7.4}$[c]	Solubility [μg mL^{-1}]/pH[d]
		a→b	b→a	(b→a)/(a→b)		
1	−4.9	7.0±0.6	9.4±0.2	1.3	1.7	>3000/6.5
24	−5.0±0.1/5.0	10.0±0.9	19.0±1.2	1.9	2.1±0.1	22±0/3.0
	−4.9±0.1/6.2					22±1/5.0
	−4.7±0.1/7.4					21±1/7.4
3	−4.9±0.0/5.0	2.2±0.2	17.6±0.4	8.0	2.0±0.0	>150/3.0
	−4.9±0.0/6.2					>150/5.0
	−4.9±0.0/7.4					>150/7.4
4a	−4.7	1.5±0.0	6.4±0.4	4.3	2.1	14±1/3.0
						13±1/5.0
						12±1/7.4
4b	n.p.	n.d.	n.d.	n.d.	<−1.5	>3000/6.6
5a	−4.6	5.3±0.6	17.5±1.3	3.3	2.3	16±2/3.0
						15±0/5.0
						17±2/7.4
5b	n.p.	0.2±0.0	0.4±0.0	1.6	−0.8	>3000/6.5
13a	−4.8±0.0/5.0	5.6±0.7	22.0±0.6	4.0	2.7±0.1	22±1/3.0
	−4.8±0.0/6.2					24±3/5.0
	−4.8±0.0/7.4					17±6/7.4
14a	−5.8±0.1/5.0	0.2±0.1	0.2±0.0	0.8	<−1.5	30±3/3.0
	−6.3±0.1/6.2					>100/5.0
	−7.4±0.1/7.4					>100/7.4
13b	−4.5±0.1/5.0	6.2±1.3	22.7±1.2	3.6	2.4±0.2	7±0/3.0
	−4.5±0.0/6.2					7±0/5.0
	−4.6±0.1/7.4					7±0/7.4
14b	−8.6±1.7/5.0	n.d.	n.d.	n.d.	−0.6±0.1	34±3/3.0
	−8.8±1.4/6.2					>200/5.0
	−8.7±1.5/7.4					>200/7.4
13c	−4.4±0.0/5.0	9.2±0.1	16.9±1.5	1.8	2.8±0.1	17±1/3.0
	−4.4±0.0/6.2					15±1/5.0
	−4.5±0.1/7.4					16±1/7.4
14c	−8.4±1.3/5.0	n.d.	n.d.	n.d.	−0.8±0.1	15±1/3.0
	−9.3±1.4/6.2					140±6/5.0
	−8.6±1.6/7.4					>200/7.4
13d	−5.4±0.0/5.0	4.2±0.7	16.4±1.2	3.9	1.8±0.1	24±0/3.0
	−5.4±0.0/6.2					24±1/5.0
	−5.4±0.0/7.4					26±1/7.4
14d	−8.5±0.6/5.0	n.d.	n.d.	n.d.	<−1.5	127±4/3.0
	−9.1±0.2/6.2					>200/5.0
	−9.2±0.4/7.4					>200/7.4
13e	−4.5±0.2/5.0	6.1±0.6	17.9±1.2	3.0	2.9±0.1	14±2/3.0
	−4.4±0.0/6.2					13±0/5.0
	−4.4±0.1/7.4					14±1/7.4
14e	−9.3±1.3/5.0	n.d.	n.d.	n.d.	−0.8±0.1	31±2/3.0
	−8.7±1.5/6.2					>200/5.0
	−8.7±1.5/7.4					>200/7.4
13f	−6.5±0.0/5.0	0.9±0.7	18.1±0.6	19.7	1.7±0.0	22±2/3.0
	−6.5±0.1/6.2					24±1/5.0
	−6.3±0.1/7.4					23±1/7.4
14f	−8.5±1.7/5.0	n.d.	n.d.	n.d.	<−1.5	35±11/3.0
	−7.3±0.3/6.2					>200/5.0
	−7.8±1.5/7.4					>200/7.4
19a	−4.9±0.0/5.0	4.4±0.1	18.8±1.7	4.3	1.9±0.1	103±8/3.0
	−4.9±0.0/6.2					100±6/5.0
	−4.9±0.1/7.4					95±5/7.4
6	−8.6±1.6/5.0	n.d.	n.d.	n.d.	<−1.5	>130/3.0
	−9.3±1.4/6.2					>130/5.0
	−8.7±1.5/7.4					>130/7.4
19b	−5.3±0.1/5.0	n.d.	n.d.	n.d.	2.4±0.1	30±0/3.0
	−5.6±0.1/6.2					29±1/5.0
	−5.1±0.2/7.4					31±1/7.4
20	−8.6±1.6/5.0	n.d.	n.d.	n.d.	−1.2±0.2	>130/3.0
	−9.3±1.4/6.2					>130/5.0
	−10/7.4					>130/7.4

Table 3. (Continued)

Compd	PAMPA $\log P_e$ [$\log 10^{-6}$ cm s^{-1}]/pH[a]	Caco-2 P_{app} [10^{-6} cm s^{-1}][b]			$\log D_{7.4}$[c]	Solubility [μg mL^{-1}]/pH[d]
		a→b	b→a	(b→a)/(a→b)		
22	−5.1 ± 0.0/5.0	n.d.	n.d.	n.d.	1.7 ± 0.1	> 130/3.0
	−5.1 ± 0.0/6.2					> 130/5.0
	−5.1 ± 0.0/7.4					> 130/7.4
23	−7.3 ± 1.8/5.0	n.d.	n.d.	n.d.	< −1.5	> 130/3.0
	−8.1 ± 2.2/6.2					> 130/5.0
	−10/7.4					> 130/7.4
30	−5.5 ± 0.0/5.0	n.d.	n.d.	n.d.	1.6 ± 0.1	> 130/3.0
	−5.5 ± 0.0/6.2					> 130/5.0
	−5.4 ± 0.1/7.4					> 130/7.4
31	−7.7 ± 1.6/5.0	n.d.	n.d.	n.d.	< −1.5	> 130/3.0
	−8.1 ± 1.3/6.2					> 130/5.0
	−10/7.4					> 130/7.4
34	−5.3 ± 0.1/5.0	n.d.	n.d.	n.d.	2.2 ± 0.1	> 130/3.0
	−5.6 ± 0.0/6.2					> 130/5.0
	−5.3 ± 0.2/7.4					> 130/7.4
35	−8.0 ± 1.3/5.0	n.d.	n.d.	n.d.	n.d.	63 ± 8/3.0
	−8.6 ± 1.6/6.2					> 130/5.0
	−10/7.4					> 130/7.4

[a] P_e = effective permeation: passive permeation through an artificial membrane was determined by parallel artificial membrane permeation assay (PAMPA); values represent the mean ± SD of quadruplicate measurements taken at three pH values (pH 5.0, 6.2, and 7.4).[33] [b] P_{app} = apparent permeability: permeation through cell monolayers was assessed by a Caco-2 assay in the absorptive (a→b) and secretory (b→a) directions in triplicate;[42] n.p. = no permeation, n.d. = not determined. [c] Distribution coefficients ($\log D$) were measured by a miniaturized shake-flask procedure at pH 7.4. [d] Kinetic solubility was measured in a 96-well format using the μSOL Explorer solubility analyzer at three pH values (pH 3.0, 5.0, and 7.4) in triplicate.

PAMPA could not be confirmed by the colorectal adenocarcinoma (Caco-2) cell permeation assay. In contrast, the methyl esters (**3**, **4a**, **5a**, **13a–f**, **19a–b**, **22**, **30**, and **34**) showed $\log D_{7.4}$ values > 1.5, that is, they are more lipophilic and hence more permeable than the corresponding acids, as shown by the PAMPA and Caco-2 permeation assay. Despite this high absorption potential, the ratios between the apparent permeability coefficients (P_{app}) in the basolateral-to-apical (b→a, secretory) and apical-to-basolateral (a→b, absorptive) directions revealed active efflux processes as an additional issue of all the assessed compounds. Moreover, the methyl esters must be readily hydrolyzed after absorption to become more polar and to be renally eliminated. Rapid metabolic turnover by the enzyme carboxylesterase was previously shown for the methyl esters **4a** and **5a**.[10e]

The different substituents at the *ortho* position of ring A (**5a**, **5b**, **13a–f**, **14a–f**; Table 3) only have a minor influence on the physicochemical properties. The addition of chloro, fluoro, methyl, trifluoromethyl, or cyclopropyl substituents slightly increases the lipophilicity of the respective acids and methyl esters, whereas methoxy and cyano substituents render the compounds more hydrophilic and less permeable. Moreover, the substituents at the *ortho* position have negligible effects on the low aqueous solubility, which is a major drawback of all methyl esters.[35] In contrast, the modifications with increased spacer length between carbohydrate and aglycone (**6**, **19a–b**, **20**, **22**, and **23**; Table 3) show higher aqueous solubility. Extending the spacer and linking it at the *ortho* or *meta* positions of the biaryl moiety disrupts the symmetry of the molecular structure, leading to increased solubility.[15,36] However, an additional chloro substituent at the 4-position (**19b**, **20**; Table 3) restores the symmetrical character of the structure, which in turn lowers the solubility of the compound. Disrupted structural symmetry might also hold true for the enhanced solubility of the biphenyls with an elongated carboxylate-bearing *para* substituent (**30**, **31**, **34**, and **35**; Table 3). The introduction of a methylene or cyclopropylene group between the biphenyl and the carboxylate moiety markedly improved the aqueous solubility of the methyl esters, whereas the absorption potential was only slightly decreased.

Summary and Conclusions

In this study, we investigated the structure–affinity relationship for *ortho* substituents on ring A of the biphenyl aglycone of the FimH antagonists **13** and **14**. The correlation between vdW volumes of these substituents and the enthalpy term clearly indicates the importance of shape complementary. This interpretation is further supported by the fact that the electronic character of the substituent [Cl in **5a** (Table 2), CF_3 in **14c** versus CH_3 in **14b**] is less important. The correlation of enthalpic improvements ($\Delta\Delta H$) with vdW volumes offers a potent tool for guiding further structural optimization.

The successful oral application using a prodrug approach was recently demonstrated with the ester **5a**.[10e] A major drawback of the biphenyl methyl esters is their insufficient solubility, which is mostly in the range of 15–35 μg mL^{-1}. As expected,[15] solubility could be substantially improved when the symmetry of the aglycone was disrupted. Thus, the solubility of **3** (> 150 μg mL^{-1}; Table 3), **19a** (100 μg mL^{-1}), and **22** (> 130 μg mL^{-1}) was improved by a factor of ~10. However, for these more flexible derivatives, the expected optimized fit leading to improved affinities in the in-docking mode could not be observed. In fact, the affinities for the members of this

family of compounds are drastically decreased, for example, compounds **20** or **23** (Table 1).

Finally, the elongation of the ester-bearing *para* substituent (Table 1; compounds **31** and **35**) did not lead to the expected additional polar interaction with Arg98. Instead, a five- to sevenfold decrease in affinity was observed. Clearly, the desolvation penalty for the guanidinium group could not be matched by the geometrically possible salt bridge with the carboxylate of the antagonists **31** and **35**.

In summary, our study confirms the earlier selection of the FimH antagonists **5a** for oral and **5b** for intravenous application. However, the methoxy derivative **13d** (Table 1) shows slightly improved pharmacokinetic properties and therefore represents an additional candidate for future in vivo studies.

Experimental Section

General methods: NMR spectra were recorded on a Bruker Avance DMX-500 (500.1 MHz) spectrometer. Assignment of ^{1}H and ^{13}C NMR spectra was achieved using 2D methods (COSY, HSQC, HMBC). Chemical shifts are expressed in ppm using residual $CHCl_3$, CHD_2OD, or HDO as references. Optical rotations were measured with a PerkinElmer Polarimeter 341. Electrospray ionization mass spectrometry (ESI-MS) data were obtained on a Waters Micromass ZQ instrument. LC–HRMS analyses were carried out using an Agilent 1100 LC equipped with a photodiode array detector and a Micromass QTOF I equipped with a 4 GHz digital time converter. Microwave-assisted reactions were carried out with a CEM Discover and Explorer. Reactions were monitored by TLC using glass plates coated with silica gel 60 F_{254} (Merck) and visualized by UV light and/or by charring with a molybdate solution (0.02 M solution of ammonium cerium sulfate dihydrate and ammonium molybdate tetrahydrate in aqueous 10% H_2SO_4). MPLC separations were carried out on a CombiFlash Companion or R_f from Teledyne Isco equipped with RediSep normal-phase or RP-18 reversed-phase flash columns. LC–MS separations were carried out on a Waters system equipped with sample manager 2767, pump 2525, PDA 2525, and Micromass ZQ. Size-exclusion chromatography was performed on Bio-Gel P-2 Gel (45–90 mm) from Bio-Rad (Reinach, Switzerland). All compounds used for biological assays are at least of 98% purity based on analytical HPLC results. Commercially available reagents were purchased from Fluka, Aldrich, Alfa Aesar or Iris Biotech (Germany). Solvents were purchased from Sigma–Aldrich (Buchs, Switzerland) or Acros Organics (Geel, Belgium) and were dried prior to use where indicated. MeOH was dried by reflux with sodium methoxide and distilled immediately before use. CH_2Cl_2 was dried by filtration over Al_2O_3 (Fluka, type 5016 A basic). Molecular sieves (4 Å) were activated in vacuo at 500 °C for 1 h immediately before use.

General procedure A for the synthesis of mannosides 10a–f and 17a–c: To an ice-cold suspension of **8**[12] (200 mg, 0.57 mmol, 1.1 equiv), phenol **9a–f** or benzyl alcohol **16a–c** (0.52 mmol, 1.0 equiv), and molecular sieves (4 Å, 600 mg) in dry CH_2Cl_2 (5 mL), $BF_3{\cdot}Et_2O$ (0.3 mL, 2.44 mmol, 4.7 equiv) was added dropwise under argon. The mixture was stirred at 0 °C for 3 h, and then at RT overnight. The reaction mixture was filtered over Celite, and the filtrate was diluted with CH_2Cl_2 (50 mL), extracted with 0.5 N $NaOH_{(aq)}$ (50 mL), H_2O (50 mL), and brine (50 mL). The organic layer was dried over Na_2SO_4 and concentrated in vacuo. The residue was purified by MPLC on silica gel (petroleum ether (PE)/EtOAc) to yield **10a–f** or **17a–c**.

General procedure B for the synthesis of mannosylated biphenyls: A Schlenk tube was charged with aryl bromide (1.0 equiv), boronic acid or boronate (1.1 equiv), $Pd(dppf)Cl_2{\cdot}CH_2Cl_2$ (0.03 equiv), K_3PO_4 (1.5 equiv) and a stirring bar. The tube was closed with a rubber septum and was evacuated and flushed with argon. This procedure was repeated once, then anhydrous DMF (2 mL) was added under a stream of argon. The mixture was degassed in an ultrasonic bath and flushed with argon for 5 min, and then stirred at 80 °C overnight. The reaction mixture was cooled to RT, diluted with EtOAc (50 mL), and washed with H_2O (50 mL) and brine (50 mL). The organic layer was dried over Na_2SO_4 and concentrated in vacuo. The residue was purified by MPLC on silica gel (PE/EtOAc) to afford biphenyls **12a–f**, **18a,b**, **21**, **29** or **33**.

General procedure C for deacetylation: To a solution of **12a–f**, **18a,b**, **21**, **29** or **33** (1.0 equiv) in dry MeOH (5 mL) was added freshly prepared 1 M NaOMe/MeOH (0.1 equiv) under argon. The mixture was stirred at RT until the reaction was complete (monitored by TLC), then neutralized with Amberlyst-15 (H^+) ion-exchange resin, filtered and concentrated in vacuo. The residue was purified by MPLC on silica gel (CH_2Cl_2/MeOH, 10:1–8:1) to afford **13a–f**, **19a,b**, **22**, **30** or **34** as white solids.

General procedure D for saponification: To a solution of **12a–e**, **18a,b**, **21**, **29** or **33** (1.0 equiv) in MeOH (5 mL) was added 1 M NaOMe/MeOH (0.1 equiv) at RT. The reaction mixture was stirred at RT for 4 h and concentrated. The residue was treated with 0.5 M $NaOH_{(aq)}$ (1 mL) for 24 h at RT. The solution was then adjusted to pH 3–4 with Amberlyst-15 (H^+), and the mixture was filtered and concentrated. The crude product was transformed into the sodium salt by passing through a small column of Dowex 50X8 (Na^+ form) ion-exchange resin. After concentration, the residue was purified by MPLC (RP-18, H_2O/MeOH, 1:0–2:1) followed by size-exclusion chromatography (P-2 gel, H_2O) to give **14a–e**, **6**, **20**, **23**, **31** or **35** as white solids after final lyophilization from H_2O.

4-Bromo-2-fluorophenyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (10a): Prepared according to general procedure A from **8**[12] and 4-bromo-2-fluorophenol (**9a**). Yield: 220 mg (74%) as white solid. R_f = 0.48 (PE/EtOAc, 2:1); $[\alpha]_D^{20}$ +83.0 (c = 0.70, EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.30 (dd, J = 2.3, 10.1 Hz, 1H, Ar-H), 7.21 (dt, J = 1.7, 8.8 Hz, 1H, Ar-H), 7.08 (t, J = 8.6 Hz, 1H, Ar-H), 5.54 (dd, J = 3.5, 10.0 Hz, 1H, H-3), 5.50 (dd, J = 1.8, 3.4 Hz, 1H, H-2), 5.46 (d, J = 1.5 Hz, 1H, H-1), 5.36 (t, J = 10.0 Hz, 1H, H-4), 4.26 (dd, J = 5.5, 12.2 Hz, 1H, H-6a), 4.17 (ddd, J = 2.1, 5.5, 10.0 Hz, 1H, H-5), 4.10 (dd, J = 2.2, 12.2 Hz, 1H, H-6b), 2.20, 2.07, 2.05, 2.04 ppm (4 s, 12H, 4 OAc); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 170.51, 169.95, 169.82, 169.76 (4 CO), 153.28 (d, J = 251.4 Hz, Ar-C), 142.64 (d, J = 11.1 Hz, Ar-C), 127.58 (d, J = 4.0 Hz, Ar-C), 120.4 (d, J = 21.5 Hz, Ar-C), 120.28 (d, J = 0.9 Hz, Ar-C), 115.73 (d, J = 8.1 Hz, Ar-C), 97.49 (C-1), 69.76 (C-5), 69.15 (C-2), 68.60 (C-3), 65.76 (C-4), 62.09 (C-6), 20.87, 20.71, 20.69, 20.67 ppm (4 $COCH_3$); elemental analysis calcd (%) for $C_{20}H_{22}BrFO_{10}$: C 46.08, H 4.25, found: C 46.11, H 4.26.

4-Bromo-2-methylphenyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (10b): Prepared according to general procedure A from **8**[12] and 4-bromo-2-methylphenol (**9b**). Yield: 254 mg (86%) as white solid. R_f = 0.60 (PE/EtOAc, 2:1); $[\alpha]_D^{20}$ +61.8 (c = 1.00, EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.31 (d, J = 1.9 Hz, 1H, Ar-H), 7.24 (dd, J = 2.3, 8.7 Hz, 1H, Ar-H), 6.97 (d, J = 8.8 Hz, 1H, Ar-H), 5.53 (dd, J = 3.4, 10.0 Hz, 1H, H-3), 5.47 (d, J = 1.7 Hz, 1H, H-1), 5.45 (dd, J = 2.0, 3.4 Hz, 1H, H-2), 5.37 (t, J = 10.0 Hz, 1H, H-4), 4.28 (dd, J = 5.6, 12.3 Hz, 1H, H-6a), 4.10–4.03 (m, 2H, H-5, H-6b), 2.27 (s, 3H, CH_3), 2.20, 2.06, 2.05, 2.04 ppm (4 s, 12H, 4 OAc); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 170.53, 170.04, 169.96, 169.73 (4 CO), 152.96, 133.78,

129.88, 129.61, 115.81, 115.23 (Ar-C), 95.91 (C-1), 69.39 (C-5), 69.38 (C-2), 68.88 (C-3), 65.76 (C-4), 62.12 (C-6), 20.88, 20.70, 20.68 (4C, 4 $COCH_3$), 16.07 ppm (CH_3); elemental analysis calcd (%) for $C_{21}H_{25}BrO_{10}$: C 48.76, H 4.87, found: C 48.84, H 4.91.

4-Bromo-2-trifluoromethyl-phenyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (10c): Prepared according to general procedure A from **8**[12] and 4-bromo-2-trifluoromethylphenol (**9c**). Yield: 260 mg (80%) as white solid. R_f = 0.50 (PE/EtOAc, 2:1); $[\alpha]_D^{20}$ +64.6 (c = 1.00, EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.74 (d, J = 2.3 Hz, 1H, Ar-H), 7.61 (dd, J = 2.4, 8.9 Hz, 1H, Ar-H), 7.15 (d, J = 8.9 Hz, 1H, Ar-H), 5.60 (d, J = 1.6 Hz, 1H, H-1), 5.51 (dd, J = 3.5, 10.1 Hz, 1H, H-3), 5.45 (dd, J = 2.0, 3.3 Hz, 1H, H-2), 5.39 (t, J = 10.1 Hz, 1H, H-4), 4.27 (dd, J = 5.3, 12.4 Hz, 1H, H-6a), 4.08–4.00 (m, 2H, H-5, H-6b), 2.21, 2.06, 2.05, 2.04 ppm (4 s, 12H, 4 OAc); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 170.41, 169.91, 169.74, 169.62 (4 CO), 152.16 (d, J = 1.7 Hz, Ar-C), 136.07 (Ar-C), 130.35 (t, J = 5.3 Hz, Ar-C), 122.30 (d, J = 271.4 Hz, CF_3), 121.72 (d, J = 31.7 Hz, Ar-C), 117.08, 114.88 (Ar-C), 95.75 (C-1), 69.96 (C-5), 69.02 (C-2), 68.45 (C-3), 65.44 (C-4), 61.95 (C-6), 20.84, 20.70, 20.67, 20.63 ppm (4 $COCH_3$); elemental analysis calcd (%) for $C_{21}H_{22}BrF_3O_{10}$: C 44.15, H 3.88, found: C 44.10, H 3.88.

4-Bromo-2-methoxyphenyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (10d): Prepared according to general procedure A from **8**[12] and 4-bromo-2-methoxyphenol (**9d**). Yield: 234 mg (77%) as white solid. R_f = 0.32 (PE/acetone, 4:1); $[\alpha]_D^{20}$ +70.3 (c = 0.70, EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.03–6.95 (m, 3H, Ar-H), 5.58 (dd, J = 3.5, 10.0 Hz, 1H, H-3), 5.52 (dd, J = 1.8, 3.4 Hz, 1H, H-2), 5.42 (d, J = 1.8 Hz, 1H, H-1), 5.34 (t, J = 10.0 Hz, 1H, H-4), 4.28–4.24 (m, 2H, H-5, H-6a), 4.10 (m, 1H, H-6b), 3.84 (s, 3H, OCH_3), 2.19, 2.07, 2.05, 2.04 ppm (4 s, 12H, 4 OAc); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 170.58, 169.98, 169.89, 169.80 (4 CO), 151.52, 143.91, 123.49, 120.37, 116.69, 115.94 (Ar-C), 97.52 (C-1), 69.45 (C-5), 69.36 (C-2), 68.80 (C-3), 66.06 (C-4), 62.27 (C-6), 56.04 (OCH_3), 20.91, 20.73, 20.71, 20.69 ppm (4 $COCH_3$); elemental analysis calcd (%) for $C_{21}H_{25}BrO_{11}$: C 47.29, H 4.72, found: C 47.20, H 4.70.

4-Bromo-2-cyclopropylphenyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (10e): Prepared according to general procedure A from **8**[12] and 4-bromo-2-cyclopropylphenol (**9e**). Yield: 235 mg (76%) as white solid. R_f = 0.30 (PE/EtOAc, 3:1); $[\alpha]_D^{20}$ +64.7 (c = 0.40, EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.20 (d, J = 8.7 Hz, 1H, Ar-H), 7.00–6.69 (m, 2H, Ar-H), 5.58 (d, J = 10.1 Hz, 1H, H-3), 5.50 (s, 2H, H-1, H-2), 5.39 (t, J = 10.1 Hz, 1H, H-4), 4.28 (dd, J = 5.4, 12.2 Hz, 1H, H-6a), 4.14–4.08 (m, 2H, H-5, H-6b), 2.21, 2.09, 2.04 (3 s, 12H, 4 OAc), 1.02 (d, J = 8.1 Hz, 2H, CH_2-*c*Pr), 0.65 ppm (d, J = 4.6 Hz, 2H, CH_2-*c*Pr); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 170.54, 170.03, 170.15, 169.75 (4 CO), 153.64, 135.64, 129.11, 128.94, 116.29, 115.79 (Ar-C), 96.15 (C-1), 69.46 (C-5), 69.39 (C-2), 68.93 (C-3), 65.78 (C-4), 62.16 (C-6), 21.07, 20.89, 20.70 (4C, 4 $COCH_3$), 9.73, 7.88, 7.82 ppm (*c*Pr); elemental analysis calcd (%) for $C_{23}H_{27}BrFO_{10}$: C 50.84, H 5.01, found: C 50.82, H 5.00.

4-Bromo-2-cyanophenyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (10f): Prepared according to general procedure A from **8**[12] and 4-bromo-2-cyanophenol (**9f**). Yield: 220 mg (73%) as white solid. R_f = 0.51 (PE/EtOAc, 2:3); $[\alpha]_D^{20}$ +54.3 (c = 0.60, EtOAc); IR (KBr): $\tilde{\nu}$ = 2232 (s, C≡N), 1749 cm^{-1} (vs, C=O); ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.73 (d, J = 2.5 Hz, 1H, Ar-H), 7.66 (dd, J = 2.5, 9.0 Hz, 1H, Ar-H), 7.15 (d, J = 9.0 Hz, 1H, Ar-H), 5.62 (d, J = 1.7 Hz, 1H, H-1), 5.56 (dd, J = 3.5, 10.0 Hz, 1H, H-3), 5.51 (dd, J = 2.0, 3.4 Hz, 1H, H-2), 5.41 (t, J = 10.0 Hz, 1H, H-4), 4.28 (dd, J = 4.9, 12.1 Hz, 1H, H-6a), 4.13–4.08 (m, 2H, H-5, H-6b), 2.21, 2.07, 2.05, 2.04 ppm (4 s, 12H, 4 OAc); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 169.37, 168.93, 168.71, 168.48 (4 CO), 155.18, 136.28, 135.00, 116.12, 114.41, 112.97, 104.62 (Ar-C, CN), 95.68 (C-1), 69.26 (C-5), 68.02 (C-2), 67.35 (C-3), 64.38 (C-4), 60.85 (C-6), 19.81, 19.67, 19.64, 19.58 ppm (4 $COCH_3$); elemental analysis calcd (%) for $C_{21}H_{22}BrNO_{10}$: C 47.74, H 4.02, N 2.65, found: C 47.78, H 4.29, N 2.67.

Methyl 4′-(2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyloxy)-3′-fluorobiphenyl-4-carboxylate (12a): Prepared according to general procedure B from **10a** (100 mg, 0.192 mmol), 4-methoxycarbonylphenylboronic acid (**11**, 38.0 mg, 0.211 mmol), Pd(dppf)Cl_2·CH_2Cl_2 (4.7 mg, 5.8 μmol) and K_3PO_4 (61.1 mg, 0.288 mmol). Yield: 83 mg (75%) as white solid. R_f = 0.26 (PE/EtOAc, 2:1); $[\alpha]_D^{20}$ +93.0 (c = 0.60, EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$): δ = 8.03–8.02 (m, 2H, Ar-H), 7.53–7.52 (m, 2H, Ar-H), 7.33 (dd, J = 2.1, 11.8 Hz, 1H, Ar-H), 7.27 (dd, J = 1.5, 8.9 Hz, 1H, Ar-H), 7.20 (t, J = 8.3 Hz, 1H, Ar-H), 5.53 (dd, J = 3.4, 10.0 Hz, 1H, H-3), 5.49–5.47 (m, 2H, H-1, H-2), 5.32 (t, J = 10.0 Hz, 1H, H-4), 4.22 (dd, J = 5.4, 12.1 Hz, 1H, H-6a), 4.17 (m, 1H, H-5), 4.05 (dd, J = 1.8, 12.1 Hz, 1H, H-6b), 3.87 (s, 3H, OCH_3), 2.15, 2.01, 1.98, 1.97 ppm (4 s, 12H, 4 OAc); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 170.54, 170.00, 169.86, 169.79, 166.82 (5 CO), 153.50 (d, J = 247.0 Hz, Ar-C), 143.56 (d, J = 1.8 Hz, Ar-C), 143.22 (d, J = 11.2 Hz, Ar-C), 136.48 (d, J = 6.7 Hz, Ar-C), 130.27, 129.29, 126.75 (5C, Ar-C), 123.16 (d, J = 3.4 Hz, Ar-C), 119.32 (Ar-C), 115.64 (d, J = 19.4 Hz, Ar-C), 97.42 (C-1), 69.71 (C-5), 69.26 (C-2), 68.70 (C-3), 65.83 (C-4), 62.10 (C-6), 52.24 (OMe), 20.91, 20.74, 20.72, 20.70 ppm (4 $COCH_3$); HRMS: m/z: calcd for $C_{28}H_{29}FNaO_{12}$ $[M+Na]^+$: 599.1535, found: 599.1536.

Methyl 4′-(2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyloxy)-3′-methylbiphenyl-4-carboxylate (12b): Prepared according to general procedure B from **10b** (100 mg, 0.193 mmol), **11** (38.2 mg, 0.212 mmol), Pd(dppf)Cl_2·CH_2Cl_2 (4.7 mg, 5.8 μmol) and K_3PO_4 (61.5 mg, 0.290 mmol). Yield: 87 mg (79%) as white solid. R_f = 0.41 (PE/EtOAc, 1:0.9); $[\alpha]_D^{20}$ +85.4 (c = 0.80, EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$): δ = 8.09–8.07 (m, 2H, Ar-H), 7.61 (m, 2H, Ar-H), 7.46 (d, J = 1.8 Hz, 1H, Ar-H), 7.40 (dd, J = 2.3, 8.5 Hz, 1H, Ar-H), 7.18 (d, J = 8.5 Hz, 1H, Ar-H), 5.61–5.58 (m, 2H, H-1, H-3), 5.50 (dd, J = 2.0, 3.5 Hz, 1H, H-2), 5.41 (t, J = 10.0 Hz, 1H, H-4), 4.31 (dd, J = 5.9, 12.8 Hz, 1H, H-6a), 4.14–4.09 (m, 2H, H-5, H-6b), 3.94 (s, 3H, OCH_3), 2.37 (s, 3H, CH_3), 2.22, 2.08, 2.05, 2.04 ppm (4 s, 12H, 4 OAc); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 170.55, 170.06, 169.98, 169.75, 167.00 (5 CO), 154.05, 144.94, 134.54, 130.10, 130.02, 128.54, 128.05, 126.66, 125.76, 114.46 (12C, Ar-C), 95.84 (C-1), 69.48 (C-5), 69.37 (C-2), 68.98 (C-3), 65.81 (C-4), 62.13 (C-6), 52.12 (OCH_3), 21.06, 20.91, 20.72, 20.70 (4 $COCH_3$), 16.40 ppm (CH_3); HRMS: m/z: calcd for $C_{29}H_{32}NaO_{12}$ $[M+Na]^+$: 595.1786, found: 595.1786; elemental analysis calcd (%) for $C_{29}H_{32}O_{12}$: C 60.84, H 5.63, found: C 60.76, H 5.80.

Methyl 4′-(2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyloxy)-3′-trifluoromethylbiphenyl-4-carboxylate (12c): Prepared according to general procedure B from **10c** (100 mg, 0.175 mmol), **11** (34.6 mg, 0.193 mmol), Pd(dppf)Cl_2·CH_2Cl_2 (4.3 mg, 5.3 μmol) and K_3PO_4 (55.7 mg, 0.263 mmol). Yield: 100 mg (91%) as white solid. R_f = 0.25 (PE/EtOAc, 2:1); $[\alpha]_D^{20}$ +43.3 (c = 1.00, EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$): δ = 8.13–8.11 (m, 2H, Ar-H), 7.87 (d, J = 2.1 Hz, 1H, Ar-H), 7.75 (dd, J = 2.2, 8.7 Hz, 1H, Ar-H), 7.63–7.61 (m, 2H, Ar-H), 7.35 (d, J = 8.7 Hz, 1H, Ar-H), 5.70 (d, J = 1.7 Hz, 1H, H-1), 5.57 (dd, J = 3.5, 10.1 Hz, 1H, H-3), 5.50 (dd, J = 2.0, 3.4 Hz, 1H, H-2), 5.43 (t, J = 10.0 Hz, 1H, H-4), 4.30 (dd, J = 5.6, 12.8 Hz, 1H, H-6a), 4.11–4.08 (m, 2H, H-5, H-6b), 3.95 (s, 3H, OCH_3), 2.24, 2.07, 2.06, 2.05 ppm (4 s, 12H, 4 OAc); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 170.45, 169.96, 169.78, 169.65, 166.76 (5 CO), 152.94, 143.37, 134.59, 130.34, 129.40, 126.79, 126.14, 115.79 (12C, Ar-C), 95.67 (C-1), 69.91 (C-5), 69.15 (C-2), 68.56 (C-3), 65.50 (C-4), 61.97 (C-6), 52.25 (OCH_3), 20.88, 20.71,

20.66 ppm (4C, 4 $COCH_3$); HRMS: *m/z*: calcd for $C_{29}H_{29}F_3NaO_{12}$ $[M+Na]^+$: 649.1503, found: 649.1503.

Methyl 4′-(2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyloxy)-3′-methoxybiphenyl-4-carboxylate (12d): Prepared according to general procedure B from **10d** (100 mg, 0.188 mmol), **11** (37.1 mg, 0.206 mmol), $Pd(dppf)Cl_2{\cdot}CH_2Cl_2$ (4.6 mg, 5.6 μmol) and K_3PO_4 (59.9 mg, 0.282 mmol). Yield: 91 mg (83%) as white solid. $R_f=0.25$ (PE/EtOAc, 1:0.9); $[\alpha]_D^{20}$ +50.7 ($c=1.40$, EtOAc); ^1H NMR (500 MHz, $CDCl_3$): $\delta=8.10–8.08$ (m, 2H, Ar-H), 7.62–7.60 (m, 2H, Ar-H), 7.19–7.13 (m, 3H, Ar-H), 5.64 (dd, $J=3.5$, 10.0 Hz, 1H, H-3), 5.58 (dd, $J=1.8$, 3.5 Hz, 1H, H-2), 5.53 (d, $J=1.7$ Hz, 1H, H-1), 5.38 (t, $J=10.0$ Hz, 1H, H-4), 4.34–4.28 (m, 2H, H-5, H-6a), 4.12 (m, 1H, H-6b), 3.94 (2 s, 6H, 2 OCH_3), 2.21, 2.08, 2.05, 2.04 ppm (4 s, 12H, 4 OAc); ^{13}C NMR (125 MHz, $CDCl_3$): $\delta=170.61$, 170.02, 169.92, 169.83, 166.94 (5 CO), 151.01, 145.06, 144.92, 136.51, 130.13, 128.86, 126.87, 119.72, 119.32, 111.60 (12C, Ar-C), 97.50 (C-1), 69.48 (C-5), 69.43 (C-2), 68.91 (C-3), 66.12 (C-4), 62.29 (C-6), 56.01 (OCH_3), 52.18 (CO_2CH_3), 20.95, 20.76, 20.74, 20.72 ppm (4 $COCH_3$); HRMS: *m/z*: calcd for $C_{29}H_{32}NaO_{13}$ $[M+Na]^+$: 611.1735, found: 611.1736.

Methyl 4′-(2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyloxy)-3′-cyclopropylbiphenyl-4-carboxylate (12e): Prepared according to general procedure B from **10e** (100 mg, 0.184 mmol), **11** (36.4 mg, 0.202 mmol), $Pd(dppf)Cl_2{\cdot}CH_2Cl_2$ (4.5 mg, 5.5 μmol) and K_3PO_4 (58.6 mg, 0.276 mmol). Yield: 60 mg (55%) as white solid. $R_f=0.48$ (PE/EtOAc, 2:1); $[\alpha]_D^{20}$ +53.0 ($c=0.70$, EtOAc); ^1H NMR (500 MHz, $CDCl_3$): $\delta=8.08–8.07$ (m, 2H, Ar-H), 7.59–7.57 (m, 2H, Ar-H), 7.37 (dd, $J=2.4$, 8.5 Hz, 1H, Ar-H), 7.19–7.17 (m, 2H, Ar-H), 5.64 (dd, $J=3.5$, 10.1 Hz, 1H, H-3), 5.61 (d, $J=1.6$ Hz, 1H, H-1), 5.54 (dd, $J=1.9$, 3.4 Hz, 1H, H-2), 5.42 (t, $J=10.1$ Hz, 1H, H-4), 4.31 (dd, $J=5.3$, 12.2 Hz, 1H, H-6a), 4.19–4.10 (m, 2H, H-5, H-6b), 3.94 (s, 3H, OCH_3), 2.22 (s, 3H, OAc), 2.17 (m, 1H, H-*c*Pr), 2.08–2.05 (m, 9H, 3 OAc), 1.06–1.05 (m, 2H, CH_2-*c*Pr), 0.74–0.73 ppm (m, 2H, CH_2-*c*Pr); ^{13}C NMR (125 MHz, $CDCl_3$): $\delta=170.55$, 170.06, 170.02, 169.75, 166.98 (5 CO), 154.76, 145.12, 134.83, 133.56, 130.08, 128.58, 126.71, 125.33, 125.06, 114.84 (12C, Ar-C), 96.04 (C-1), 69.49 (C-5), 69.42 (C-2), 69.02 (C-3), 65.81 (C-4), 62.15 (C-6), 52.12 (OCH_3), 20.91, 20.71 (4C, 4 $COCH_3$), 9.78, 7.58 ppm (3C, *c*Pr); elemental analysis calcd (%) for $C_{29}H_{32}O_{12}$: C 62.20, H 5.72, found: C 62.00, H 5.86.

Methyl 4′-(2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyloxy)-3′-cyanobiphenyl-4-carboxylate (12f): Prepared according to general procedure B from **10f** (100 mg, 0.189 mmol), **11** (37.5 mg, 0.208 mmol), $Pd(dppf)Cl_2{\cdot}CH_2Cl_2$ (4.6 mg, 5.7 μmol) and K_3PO_4 (60.2 mg, 0.284 mmol). Yield: 92 mg (84%) as white solid. $R_f=0.18$ (PE/EtOAc, 2:1); $[\alpha]_D^{20}$ +61.4 ($c=0.80$, EtOAc); ^1H NMR (500 MHz, $CDCl_3$): $\delta=8.06–8.05$ (m, 2H, Ar-H), 7.80 (d, $J=2.3$ Hz, 1H, Ar-H), 7.72 (dd, $J=2.3$, 8.8 Hz, 1H, Ar-H), 7.53–7.51 (m, 2H, Ar-H), 7.28 (d, $J=8.8$ Hz, 1H, Ar-H), 5.64 (d, $J=1.7$ Hz, 1H, H-1), 5.55 (dd, $J=3.5$, 10.0 Hz, 1H, H-3), 5.49 (dd, $J=1.9$, 3.4 Hz, 1H, H-2), 5.37 (t, $J=10.0$ Hz, 1H, H-4), 4.24 (dd, $J=5.0$, 12.4 Hz, 1H, H-6a), 4.12 (ddd, $J=2.2$, 4.9, 10.0 Hz, 1H, H-5), 4.05 (dd, $J=2.2$, 12.4 Hz, 1H, H-6b), 3.88 (s, 3H, OCH_3), 2.16, 2.01, 1.99, 1.98 ppm (4 s, 12H, 4 OAc); ^{13}C NMR (125 MHz, $CDCl_3$): $\delta=170.45$, 170.01, 169.78, 169.54, 166.65 (5 CO), 156.84, 142.44, 135.67, 132.36, 129.77, 126.76, 115.99, 115.18, 104.47 (13C, Ar-C, CN), 96.63 (C-1), 70.24 (C-5), 69.17 (C-2), 68.49 (C-3), 65.48 (C-4), 60.85 (C-6), 20.88, 20.73, 20.71, 20.64 ppm (4 $COCH_3$); HRMS: *m/z*: calcd for $C_{29}H_{29}NNaO_{12}$ $[M+Na]^+$: 606.1582, found: 606.1583.

Methyl 3′-fluoro-4′-(α-D-mannopyranosyloxy)biphenyl-4-carboxylate (13a): Prepared according to general procedure C from **12a** (33 mg, 0.057 mmol). Yield: 15 mg (65%). $[\alpha]_D^{20}$ +114.3 ($c=0.30$, MeOH); ^1H NMR (500 MHz, CD_3OD): $\delta=7.98–7.97$ (m, 2H, Ar-H), 7.63–7.61 (m, 2H, Ar-H), 7.42–7.36 (m, 3H, Ar-H), 5.45 (d, $J=1.7$ Hz, 1H, H-1), 3.99 (dd, $J=1.9$, 3.4 Hz, 1H, H-2), 3.82–3.84 (m, 4H, H-3, OCH_3), 3.71–3.56 ppm (m, 4H, H-4, H-5, H-6); ^{13}C NMR (125 MHz, CD_3OD): $\delta=168.34$ (CO), 154.75 (d, $J=243.8$ Hz, Ar-C), 145.6 (2C, Ar-C), 136.37 (d, $J=6.9$ Hz, Ar-C), 130.20, 129.20, 127.80, 124.33, 120.33 (7C, Ar-C), 116.00 (d, $J=20.0$ Hz, Ar-C), 101.40 (C-1), 75.97 (C-5), 72.31 (C-3), 71.82 (C-2), 68.18 (C-4), 62.65 (C-6), 52.65 ppm (OCH_3); HRMS: *m/z*: calcd for $C_{20}H_{21}FNaO_8$ $[M+Na]^+$: 431.1113, found: 431.1112.

Methyl 4′-(α-D-mannopyranosyloxy)-3′-methylbiphenyl-4-carboxylate (13b): Prepared according to general procedure C from **12b** (31 mg, 0.054 mmol). Yield: 16 mg (73%). $[\alpha]_D^{20}$ +110.5 ($c=0.35$, MeOH); ^1H NMR (500 MHz, CD_3OD): $\delta=7.96–7.94$ (m, 2H, Ar-H), 7.60–7.58 (m, 2H, Ar-H), 7.40–7.37 (m, 2H, Ar-H), 7.22 (d, $J=8.5$ Hz, 1H, Ar-H), 5.47 (d, $J=1.6$ Hz, 1H, H-1), 3.97 (dd, $J=1.9$, 3.4 Hz, 1H, H-2), 3.87 (dd, $J=3.4$, 9.5 Hz, 1H, H-3), 3.82 (s, 3H, OMe), 3.67–3.52 (m, 3H, H-4, H-6), 3.46 (m, 1H, H-5), 2.21 ppm (s, 3H, Me); ^{13}C NMR (125 MHz, CD_3OD): $\delta=168.56$ (CO), 156.20, 146.86, 134.70, 131.07, 130.07, 130.54, 129.45, 128.92, 127.63, 126.85, 115.83 (12C, Ar-C), 99.76 (C-1), 75.55 (C-5), 72.64 (C-3), 72.11 (C-2), 68.31 (C-4), 62.68 (C-6), 52.59 ppm (OCH_3), 16.54 (CH_3); HRMS: *m/z*: calcd for $C_{21}H_{24}NaO_8$ $[M+Na]^+$: 427.1363, found: 427.1370.

Methyl 3′-trifluoromethyl-4′-(α-D-mannopyranosyloxy)biphenyl-4-carboxylate (13c): Prepared according to general procedure C from **12c** (30 mg, 0.048 mmol). Yield: 14 mg (64%). $[\alpha]_D^{20}$ +113.1 ($c=0.40$, MeOH); ^1H NMR (500 MHz, CD_3OD): $\delta=8.11–8.10$ (m, 2H, Ar-H), 7.92–7.90 (m, 2H, Ar-H), 7.75–7.73 (m, 2H, Ar-H), 7.63 (d, $J=8.4$ Hz, 1H, Ar-H), 5.69 (d, $J=1.5$ Hz, 1H, H-1), 4.09 (dd, $J=1.8$, 3.3 Hz, 1H, H-2), 3.98–3.94 (m, 4H, H-3, OMe), 3.79–3.73 (m, 3H, H-4, H-6), 3.61 ppm (ddd, $J=2.3$, 5.7, 9.6 Hz, 1H, H-5); ^{13}C NMR (125 MHz, CD_3OD): $\delta=168.29$ (CO), 155.54, 145.13, 134.74, 133.45, 131.36, 131.29, 130.32, 127.91, 127.85, 126.44, 117.80 (Ar-C), 100.27 (C-1), 76.13 (C-5), 72.24 (C-3), 71.74 (C-2), 68.09 (C-4), 62.67 ppm (C-6), 52.69 (OMe); HRMS: *m/z*: calcd for $C_{21}H_{21}F_3NaO_8$ $[M+Na]^+$: 481.1081, found: 481.1082.

Methyl 4′-(α-D-mannopyranosyloxy)-3′-methoxybiphenyl-4-carboxylate (13d): Prepared according to general procedure C from **12d** (32 mg, 0.055 mmol). Yield: 12 mg (52%). $[\alpha]_D^{20}$ +133.1 ($c=0.20$, MeOH); ^1H NMR (500 MHz, CD_3OD): $\delta=7.97–7.96$ (m, 2H, Ar-H), 7.63–7.61 (m, 2H, Ar-H), 7.21–7.11 (m, 3H, Ar-H), 5.37 (d, $J=1.7$ Hz, 1H, H-1), 4.00 (dd, $J=1.8$, 3.4 Hz, 1H, H-2), 3.86 (dd, $J=3.5$, 8.8 Hz, 1H, H-3), 3.82 (s, 6H, 2 CH_3), 3.70–3.63 ppm (m, 4H, H-4, H-5, H-6); ^{13}C NMR (125 MHz, CD_3OD): $\delta=168.50$ (CO), 152.33, 147.40, 146.83, 136.56, 131.08, 129.76, 127.87, 120.86, 120.10, 112.54 (Ar-C), 101.51 (C-1), 75.66 (C-5), 72.40 (C-2), 72.00 (C-3), 68.34 (C-4), 62.70 (C-6), 56.61 (OMe), 52.63 ppm (OMe); HRMS: *m/z*: calcd for $C_{21}H_{24}NaO_9$ $[M+Na]^+$: 443.1313, found: 443.1315.

Methyl 3′-cyclopropyl-4′-(α-D-mannopyranosyloxy)biphenyl-4-carboxylate (13e): Prepared according to general procedure C from **12e** (21 mg, 0.035 mmol). Yield: 10 mg (67%). $[\alpha]_D^{20}$ +101.6 ($c=0.24$, MeOH); ^1H NMR (500 MHz, CD_3OD): $\delta=8.07–8.05$ (m, 2H, Ar-H), 7.68–7.67 (m, 2H, Ar-H), 7.46 (dd, $J=2.4$, 8.5 Hz, 1H, Ar-H), 7.33 (d, $J=8.5$ Hz, 1H, Ar-H), 7.21 (d, $J=2.4$ Hz, 1H, Ar-H), 5.60 (d, $J=1.7$ Hz, 1H, H-1), 4.13 (dd, $J=1.9$, 3.3 Hz, 1H, H-2), 4.02 (dd, $J=3.4$, 9.5 Hz, 1H, H-3), 3.93 (s, 3H, OMe), 3.81–3.74 (m, 3H, H-4, H-6), 3.69 (m, 1H, H-5), 2.19 (m, 1H, H-*c*Pr), 1.01–0.99 (m, 2H, CH_2-*c*Pr), 0.76–0.74 ppm (m, 2H, CH_2-*c*Pr); ^{13}C NMR (125 MHz, CD_3OD): $\delta=168.54$ (CO), 156.92, 146.98, 135.00, 134.59, 131.07, 127.34, 127.67, 126.39, 125.34, 116.29 (12C, Ar-C), 100.14 (C-1), 75.61 (C-5), 72.64 (C-3), 72.14 (C-2), 68.33 (C-4), 62.71 (C-6), 52.60 (OCH_3), 10.93,

8.06 ppm (3C, *c*Pr); HRMS: *m/z*: calcd for $C_{23}H_{26}NaO_8$ $[M+Na]^+$: 453.1520, found: 453.1519.

Methyl 3′-cyano-4′-(α-D-mannopyranosyloxy)biphenyl-4-carboxylate (13 f): Prepared according to general procedure C from **12 f** (37 mg, 0.063 mmol). Yield: 19 mg (73%). $[\alpha]_D^{20}$ +101.1 (*c* = 0.30, MeOH); ^{1}H NMR (500 MHz, CD_3OD): δ = 8.00–7.99 (m, 2H, Ar-H), 7.90–7.85 (m, 2H, Ar-H), 7.65–7.63 (m, 2H, Ar-H), 7.50 (d, *J* = 8.8 Hz, 1H, Ar-H), 5.63 (s, 1H, H-1), 4.03 (m, 1H, H-2), 3.91 (dd, *J* = 2.8, 9.4 Hz, 1H, H-3), 3.83 (s, 3H, OMe), 3.69–3.60 (m, 3H, H-4, H-6), 3.50 ppm (m, 1H, H-5); ^{13}C NMR (125 MHz, CD_3OD): δ = 168.22 (CO), 159.29, 144.38, 135.61, 134.50, 133.08, 131.31, 130.56, 127.87, 117.36, 116.75, 104.35 (13C, Ar-C, CN), 100.62 (C-1), 76.39 (C-5), 72.27 (C-2), 71.62 (C-3), 68.07 (C-4), 62.64 (C-6), 52.71 ppm (OMe); HRMS: *m/z*: calcd for $C_{21}H_{21}NNaO_8$ $[M+Na]^+$: 438.1159, found: 438.1162.

Sodium 3′-fluoro-4′-(α-D-mannopyranosyloxy)biphenyl-4-carboxylate (14 a): Prepared according to general procedure D from **12 a** (50 mg, 0.087 mmol). Yield: 21 mg (58%). $[\alpha]_D^{20}$ +112.7 (*c* = 0.40, MeOH); ^{1}H NMR (500 MHz, D_2O): δ = 7.78–7.77 (m, 2H, Ar-H), 7.46–7.45 (m, 2H, Ar-H), 7.30–7.15 (m, 3H, Ar-H), 5.43 (s, 1H, H-1), 4.07 (s, 1H, H-2), 3.93 (d, *J* = 3.3 Hz, 1H, H-3), 3.68–3.62 ppm (m, 4H, H-4, H-5, H-6); ^{13}C NMR (125 MHz, D_2O): δ = 175.19 (CO), 153.02 (d, *J* = 242.6 Hz, Ar-C), 142.52 (d, *J* = 10.8 Hz, Ar-C), 141.23 (Ar-C), 135.53 (d, *J* = 6.4 Hz, Ar-C), 135.07, 129.43, 126.25, 126.01, 122.96, 119.13 (Ar-C), 114.83 (d, *J* = 19.4 Hz, Ar-C), 99.32 (C-1), 73.65 (C-5), 70.23 (C-3), 69.67 (C-2), 66.35 (C-4), 60.52 ppm (C-6); HRMS: *m/z*: calcd for $C_{19}H_{19}FNaO_8$ $[M+Na]^+$: 417.0956, found: 417.0957.

Sodium 4′-(α-D-mannopyranosyloxy)-3′-methylbiphenyl-4-carboxylate (14 b): Prepared according to general procedure D from **12 b** (46 mg, 0.081 mmol). Yield: 5 mg (15%). $[\alpha]_D^{20}$ +85.7 (*c* = 0.20, MeOH); ^{1}H NMR (500 MHz, D_2O): δ = 7.78–7.76 (m, 2H, Ar-H), 7.53–7.52 (m, 2H, Ar-H), 7.43–7.37 (m, 2H, Ar-H), 7.10 (d, *J* = 8.6 Hz, 1H, Ar-H), 5.52 (d, *J* = 1.6 Hz, 1H, H-1), 4.07 (dd, *J* = 1.9, 3.4 Hz, 1H, H-2), 3.95 (dd, *J* = 3.5, 9.6 Hz, 1H, H-3), 3.63–3.50 (m, 4H, H-4, H-5, H-6), 2.14 ppm (s, 3H, CH_3); ^{13}C NMR (125 MHz, D_2O): δ = 153.33, 142.57, 134.59, 133.97, 129.47, 128.42, 126.25, 125.43, 114.99 (12C, Ar-C), 97.46 (C-1), 73.39 (C-5), 70.54 (C-3), 69.88 (C-2), 66.53 (C-4), 60.60 (C-6), 15.31 ppm (CH_3); HRMS: *m/z*: calcd for $C_{20}H_{22}NaO_8$ $[M+Na]^+$: 413.1207, found: 413.1209.

Sodium 3′-trifluoromethyl-4′-(α-D-mannopyranosyloxy)biphenyl-4-carboxylate (14 c): Prepared according to general procedure D from **12 c** (45 mg, 0.072 mmol). Yield: 25 mg (74%). $[\alpha]_D^{20}$ +94.2 (*c* = 0.30, MeOH); ^{1}H NMR (500 MHz, D_2O): δ = 7.83–7.81(m, 3H, Ar-H), 7.75 (d, *J* = 8.7 Hz, 1H, Ar-H), 7.57–7.55 (m, 2H, Ar-H), 7.31 (d, *J* = 8.8 Hz, 1H, Ar-H), 5.64 (s, 1H, H-1), 4.09 (d, *J* = 1.5 Hz, 1H, H-2), 3.94 (dd, *J* = 3.4, 9.7 Hz, 1H, H-3), 3.67–3.60 (m, 3H, H-4, H-6), 3.54 ppm (m, 1H, H-5); ^{13}C NMR (125 MHz, D_2O): δ = 175.25 (CO), 152.40, 141.31, 135.09, 133.53, 131.93, 129.46, 126.34, 125.59, 115.86 (12C, Ar-C), 97.20 (C-1), 73.68 (C-5), 70.19 (C-3), 69.58 (C-2), 66.36 (C-4), 60.55 ppm (C-6); HRMS: *m/z*: calcd for $C_{20}H_{19}F_3NaO_8$ $[M+Na]^+$: 467.0924, found: 467.0923.

Sodium 4′-(α-D-mannopyranosyloxy)-3′-methoxybiphenyl-4-carboxylate (14 d): Prepared according to general procedure D from **12 d** (47 mg, 0.080 mmol). Yield: 10 mg (29%). $[\alpha]_D^{20}$ +115.1 (*c* = 0.30, MeOH); ^{1}H NMR (500 MHz, D_2O): δ = 7.81–7.79 (m, 2H, Ar-H), 7.54–7.53 (m, 2H, Ar-H), 7.19–7.11 (m, 3H, Ar-H), 5.43 (d, *J* = 1.6 Hz, 1H, H-1), 4.10 (dd, *J* = 1.8, 3.5 Hz, 1H, H-2), 3.96 (dd, *J* = 3.5, 9.0 Hz, 1H, H-3), 3.78 (s, 3H, OCH_3), 3.70–3.62 ppm (m, 4H, H-4, H-5, H-6); ^{13}C NMR (125 MHz, D_2O): δ = 175.24 (CO), 149.53, 144.24, 142.42, 135.59, 134.75, 129.40, 126.41, 119.86, 118.03, 111.44 (12C, Ar-C), 99.23 (C-1), 73.53 (C-5), 70.32 (C-3), 69.78 (C-2), 66.40 (C-4), 60.54 (C-6), 55.81 ppm (OCH_3); HRMS: *m/z*: calcd for $C_{20}H_{22}NaO_9$ $[M+Na]^+$: 429.1156, found: 429.1154.

Sodium 3′-cyclopropyl-4′-(α-D-mannopyranosyloxy)biphenyl-4-carboxylate (14 e): Prepared according to general procedure D from **12 e** (28 mg, 0.047 mmol). Yield: 6 mg (26%). $[\alpha]_D^{20}$ +149.8 (*c* = 0.20, MeOH); ^{1}H NMR (500 MHz, D_2O): δ = 7.79–7.77 (m, 2H, Ar-H), 7.48–7.46 (m, 2H, Ar-H), 7.30 (d, *J* = 7.8 Hz, 1H, Ar-H), 7.07–7.05 (m, 2H, Ar-H), 5.52 (s, 1H, H-1), 4.10 (m, 1H, H-2), 3.98 (dd, *J* = 3.4, 9.5 Hz, 1H, H-3), 3.69–3.62 (m, 4H, H-4, H-5, H-6), 1.99 (m, 1H, H-*c*Pr), 0.86–0.84 (m, 2H, CH_2-*c*Pr), 0.58–0.56 ppm (m, 2H, CH_2-*c*Pr); ^{13}C NMR (125 MHz, D_2O): δ = 175.34 (CO), 153.82, 142.58, 134.57, 134.34, 133.74, 129.38, 126.26, 125.01, 124.01, 115.47 (12C, Ar-C), 97.88 (C-1), 73.47 (C-5), 70.55 (C-3), 69.93 (C-2), 66.46 (C-4), 60.57 (C-6), 9.16, 7.26, 7.06 ppm (*c*Pr); HRMS: *m/z*: calcd for $C_{22}H_{24}ONaO_8$ $[M+Na]^+$: 439.1363, found: 439.1363.

Sodium 3′-cyano-4′-(α-D-mannopyranosyloxy)biphenyl-4-carboxylate (14 f): A two-neck flask was charged with **10 f** (150 mg, 0.28 mmol), 4-carboxybenzene boronic acid pinacol ester (**15**) (77 mg, 0.31 mmol), Pd(dppf)Cl_2· CH_2Cl_2 (7 mg, 0.008 mmol), K_3PO_4 (89 mg, 0.42 mmol) and a stirring bar. The flask was evacuated and flushed with argon, then anhydrous DMF (2 mL) was added under a stream of argon. The mixture was degassed in an ultrasonic bath and flushed with argon for 5 min, and then stirred at 80 °C overnight. The reaction mixture was cooled to RT, diluted with EtOAc (50 mL), and washed with H_2O (50 mL) and brine (50 mL). The organic layer was dried over Na_2SO_4 and concentrated in vacuo. The residue was purified by MPLC on silica gel (CH_2Cl_2/MeOH, 10:1–8:1) to afford the biphenyl intermediate (162 mg). The intermediate was dissolved in dry MeOH (4 mL) and treated with freshly prepared 1 M NaOMe/MeOH (28 μL) for 4 h at RT. The reaction mixture was neutralized with Amberlyst-15 (H^+), filtered and concentrated. The crude product was transformed into the sodium salt by passing through a small column of Dowex 50X8 (Na^+ form) ion-exchange resin. After concentration the residue was purified by MPLC (RP-18, H_2O) followed by size-exclusion chromatography (P-2 gel, H_2O) to give **14 f** (19 mg, 17%) as a white solid after final lyophilization from H_2O. $[\alpha]_D^{20}$ +75.3 (*c* = 0.20, MeOH); ^{1}H NMR (500 MHz, D_2O): δ = 7.86–7.79 (m, 4H, Ar-H), 7.53–7.52 (m, 2H, Ar-H), 7.31 (d, *J* = 8.9 Hz, 1H, Ar-H), 5.64 (d, *J* = 1.9 Hz, 1H, H-1), 4.11 (dd, *J* = 1.9, 3.4 Hz, 1H, H-2), 4.00 (dd, *J* = 3.5, 9.7 Hz, 1H, H-3), 3.73–3.65 (m, 3H, H-4, H-6), 3.58 ppm (ddd, *J* = 2.4, 5.5, 9.9 Hz, 1H, H-5); ^{13}C NMR (125 MHz, D_2O): δ = 175.12 (CO), 156.82, 140.37, 134.39, 133.56, 131.83, 129.58, 126.25, 116.82, 115.78, 102.08 (13C, Ar-C, CN), 98.09 (C-1), 73.97 (C-5), 70.29 (C-3), 69.54 (C-2), 66.36 (C-4), 60.56 ppm (C-6); HRMS: *m/z*: calcd for $C_{21}H_{21}NNaO_8$ $[M+Na]^+$: 424.1003, found: 424.1003.

3-Bromobenzyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (17 a): Prepared according to general procedure A from **8**[12] and 3-bromobenzyl alcohol (**16 a**). Yield: 100 mg (34%) as colorless oil. R_f = 0.43 (PE/EtOAc, 2:1); $[\alpha]_D^{20}$ +42.0 (*c* = 1.40, EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.48–7.46, 7.30–7.24 (m, 4H, Ar-H), 5.37 (dd, 1H, *J* = 3.4, 10.1 Hz, H-3), 5.33–5.29 (m, 2H, H-2, H-4), 4.88 (d, 1H, *J* = 1.3 Hz, H-1), 4.68, 4.54 (A, B of AB, *J* = 12.1 Hz, 2H, CH_2Ar), 4.29 (dd, 1H, *J* = 5.2, 12.3 Hz, H-6a), 4.07 (dd, 1H, *J* = 2.3, 12.3 Hz, H-6b), 3.99 (ddd, 1H, *J* = 2.4, 5.2, 9.9 Hz, H-5), 2.15, 2.13, 2.05, 2.00 ppm (4 s, 12H, 4 OAc); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 170.59, 169.98, 169.87, 169.69 (4 CO), 138.49, 131.34, 131.09, 130.24, 126.66, 122.57 (Ar-C), 96.83 (C-1), 69.43, 69.02, 68.90, 68.78 (C-2, C-3, C-5, CH_2Ar), 66.03 (C-4), 62.36 (C-6), 20.86, 20.76, 20.68 ppm (4C, COCH_3); ESI-MS: *m/z*: calcd for $C_{21}H_{25}BrNaO_{10}$ $[M+Na]^+$: 539.05, found: 539.14.

5-Bromo-2-chlorobenzyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (17b): Prepared according to general procedure A from **8**[12] and 5-bromo-2-chlorobenzyl alcohol (**16b**). Yield: 152 mg (48%) as a white solid. R_f = 0.56 (PE/EtOAc, 2:1); $[\alpha]_D^{20}$ +48.0 (*c* = 1.50, EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.48 (t, *J* = 1.8 Hz, 1H, Ar-H), 7.38 (s, 1H, Ar-H), 7.35 (d, *J* = 1.8 Hz, 1H, Ar), 5.33 (m, 3H, H-2, H-3, H-4), 4.88 (d, *J* = 1.5 Hz, 1H, H-1), 4.65, 4.51 (A, B of AB, *J* = 12.3 Hz, 2H, CH_2Ar), 4.30 (dd, *J* = 5.3, 12.3 Hz, 1H, H-6a), 4.09 (dd, *J* = 2.4, 12.3 Hz, 1H, H-6b), 3.98 (ddd, *J* = 2.4, 5.2, 9.7 Hz, 1H, H-5), 2.16, 2.13, 2.05, 2.01 ppm (4 s, 12H, 4 OAc); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 170.58, 169.98, 169.89, 169.69 (4 CO), 139.77, 135.35, 129.25, 126.85, 122.91 (6C, Ar-C), 96.96 (C-1), 69.33, 68.93, 68.24 (4C, C-2, C-3, C-5, CH_2Ar), 65.98 (C-4), 62.38 (C-6), 20.86, 20.77, 20.68 ppm (4C, 4COCH_3); ESI-MS: *m/z*: calcd for $C_{21}H_{24}BrClNaO_{10}$ $[M+Na]^+$: 573.01, found: 573.06.

2-Bromobenzyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (17c): Prepared according to general procedure A from **8**[12] and 2-bromobenzyl alcohol (**16c**). Yield: 140 mg (47%) as a white solid. R_f = 0.55 (petrol ether/EtOAc, 2:1); $[\alpha]_D^{20}$ +44.6 (*c* = 2.10, EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.57 (dd, *J* = 1.0, 8.0 Hz, 1H, Ar-H), 7.47 (dd, *J* = 1.4, 7.6 Hz, 1H, Ar-H), 7.35 (td, *J* = 1.1, 7.5 Hz, 1H, Ar-H), 7.20 (td, *J* = 1.7, 7.9 Hz, 1H, Ar-H), 5.41 (dd, *J* = 3.5, 10.0 Hz, 1H, H-3), 5.35 (dd, *J* = 1.8, 3.5 Hz, 1H, H-2), 5.31 (t, *J* = 9.9 Hz, 1H, H-4), 4.98 (d, *J* = 1.6 Hz, 1H, H-1), 4.83, 4.61 (A, B of AB, *J* = 12.7 Hz, 2H, CH_2Ar), 4.29 (dd, *J* = 5.8, 12.6 Hz, 1H, H-6a), 4.10–4.06 (m, 2H, H-6b, H-5), 2.17, 2.12, 2.04, 2.00 ppm (4 s, 12H, 4 OAc); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 170.64, 170.02, 169.88, 169.72 (4 CO), 135.77, 132.69, 129.58, 129.49, 127.64, 122.96 (Ar-C), 97.33 (C-1), 69.48, 69.30, 69.10, 68.84 (C-2, C-3, C-5, CH_2Ar), 66.05 (C-4), 62.35 (C-6), 20.88, 20.76, 20.69 ppm (4C, 4COCH_3); ESI-MS: *m/z*: calcd for $C_{21}H_{25}BrNaO_{10}$ $[M+Na]^+$: 539.05, found: 539.14.

Methyl 3′-[(2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyloxy)methyl]biphenyl-4-carboxylate (18a): Prepared according to general procedure B from **17a** (87.0 mg, 0.167 mmol), **11** (33.1 mg, 0.184 mmol), Pd(dppf)Cl_2·CH_2Cl_2 (4.1 mg, 5.0 μmol) and K_3PO_4 (53.2 mg, 0.251 mmol). Yield: 70 mg (73%) as colorless oil. R_f = 0.30 (PE/EtOAc, 2:1); $[\alpha]_D^{20}$ +41.2 (*c* = 1.00, EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$): δ = 8.13–8.11 (m, 2H, Ar-H), 7.68–7.67 (m, 2H, Ar-H), 7.60–7.58 (m, 2H, Ar-H), 7.48 (t, *J* = 4.7 Hz, 1H, Ar-H), 7.39 (d, *J* = 7.7 Hz, 1H, Ar-H), 5.41 (dd, *J* = 3.4, 10.0 Hz, 1H, H-3), 5.33–5.30 (m, 2H, H-2, H-4), 4.94 (d, *J* = 1.5 Hz, 1H, H-1), 4.79, 4.64 (A, B of AB, *J* = 12.0 Hz, 2H, CH_2Ar), 4.30 (dd, *J* = 5.0, 12.1 Hz, 1H, H-6a), 4.09–4.03 (m, 2H, H-6b, H-5), 3.94 (s, 3H, OMe), 2.15, 2.11, 2.04, 2.00 ppm (4 s, 12H, 4 OAc); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 170.64, 170.03, 169.91, 169.73, 166.94 (5 CO), 145.11, 140.41, 136.97, 130.15, 129.27, 129.09, 127.94, 127.22, 127.11 (12C, Ar-C), 96.76 (C-1), 69.57, 69.09, 68.94, 66.12 (C-2, C-3, C-5, CH_2Ar), 62.40 (C-4), 60.38 (C-6), 52.15 (OMe), 20.89, 20.77, 20.69 ppm (4C, 4COCH_3); ESI-MS: *m/z*: calcd for $C_{29}H_{32}NaO_{12}$ $[M+Na]^+$: 595.18, found: 595.21.

Methyl 3′-[(2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyloxy)methyl]-4′-chlorobiphenyl-4-carboxylate (18b): Prepared according to general procedure B from **17b** (143 mg, 0.260 mmol), **11** (51.5 mg, 0.286 mmol), Pd(dppf)Cl_2·CH_2Cl_2 (6.4 mg, 7.8 μmol) and K_3PO_4 (82.8 mg, 0.390 mmol). Yield: 133 mg (84%) as colorless oil. R_f = 0.30 (PE/EtOAc, 2:1); $[\alpha]_D^{20}$ +45.9 (*c* = 1.20, EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$): δ = 8.13–8.11 (m, 2H, Ar-H), 7.65–7.64 (m, 2H, Ar-H), 7.57 (t, *J* = 1.8 Hz, 1H, Ar-H), 7.47 (s, 1H, Ar-H), 7.37 (s, 1H, Ar-H), 5.40 (dd, *J* = 3.4, 10.1 Hz, 1H, H-3), 5.33–5.29 (m, 2H, H-2, H-4), 4.93 (d, *J* = 1.4 Hz, 1H, H-1), 4.76, 4.61 (A, B of AB, *J* = 12.1 Hz, 2H, CH_2Ar), 4.31 (dd, *J* = 5.2, 12.3 Hz, 1H, H-6a), 4.11 (dd, *J* = 2.3, 12.3 Hz, 1H, H-6b), 4.03 (ddd, *J* = 2.4, 5.2, 9.9 Hz, 1H, H-5), 3.95 (s, 3H, OMe), 2.16, 2.12, 2.05, 2.00 ppm (4 s, 12H, 4OAc); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 170.61, 170.02, 169.90, 169.72, 166.75 (5 CO), 143.68, 142.16, 138.79, 135.11, 130.26, 129.68, 127.14, 125.14 (12C, Ar-C), 96.85 (C-1), 68.99, 68.89, 68.85, 66.07 (C-2, C-3, C-5, CH_2Ar), 62.42 (C-4), 60.39 (C-6), 52.23 (OMe), 20.89, 20.78, 20.71, 20.69 ppm (4COCH_3); ESI-MS: *m/z*: calcd for $C_{29}H_{31}ClNaO_{12}$ $[M+Na]^+$: 629.14, found: 629.10.

Methyl 2′-[(2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyloxy)methyl]biphenyl-4-carboxylate (21): Prepared according to general procedure B from **17c** (115 mg, 0.223 mmol), **11** (44.1 mg, 0.245 mmol), Pd(dppf)Cl_2·CH_2Cl_2 (5.5 mg, 6.7 μmol) and K_3PO_4 (71.0 mg, 0.335 mmol). Yield: 120 mg (94%) as colorless oil. R_f = 0.41 (PE/EtOAc, 2:1); $[\alpha]_D^{20}$ +38.3 (*c* = 2.00, EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$): δ = 8.11–8.10 (m, 2H, Ar-H), 7.51–7.48 (m, 1H, Ar-H), 7.45–7.41 (m, 4H, Ar-H), 7.29 (m, 1H, Ar-H), 5.27–5.21 (m, 2H, H-3, H-4), 5.19 (dd, *J* = 1.9, 3.3 Hz, 1H, H-2), 4.77 (d, *J* = 1.4 Hz, 1H, H-1), 4.67, 4.34 (A, B of AB, *J* = 11.3 Hz, 2H, CH_2Ar), 4.13 (dd, *J* = 5.2, 12.5 Hz, 1H, H-6a), 3.94 (s, 3H, OMe), 3.90 (dd, *J* = 2.2, 12.3 Hz, 1H, H-6a), 3.52 (ddd, *J* = 2.2, 5.1, 9.3 Hz, 1H, H-5), 2.13, 2.05, 2.04, 1.99 ppm (4 s, 12H, 4 OAc); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 170.52, 169.95, 169.82, 169.74, 166.77 (5 CO), 145.48, 141.44, 133.46, 129.99, 129.91, 129.58, 129.22, 128.52, 128.21 (12C, Ar-C), 97.20 (C-1), 69.47 (C-2), 68.98 (C-3), 68.48 (C-5), 68.13 (CH_2Ar), 65.88 (C-4), 62.15 (C-6), 52.18 (OMe), 20.85, 20.66, 20.62 ppm (4C, 4COCH_3); ESI-MS: *m/z*: calcd for $C_{29}H_{32}NaO_{12}$ $[M+Na]^+$: 595.18, found: 595.21.

Methyl 3′-[(α-D-mannopyranosyloxy)methyl]biphenyl-4-carboxylate (19a): Prepared according to general procedure C from **18a** (24 mg, 0.042 mmol). Yield: 11 mg (65%). R_f = 0.40 (CH_2Cl_2/MeOH, 8:1); $[\alpha]_D^{20}$ +68.0 (*c* = 0.34, MeOH); ^{1}H NMR (500 MHz, CD_3OD): δ = 8.11–8.09 (m, 2H, Ar-H), 7.77–7.75 (m, 2H, Ar-H), 7.70 (s, 1H, Ar-H), 7.63 (d, *J* = 7.6 Hz, 1H, Ar-H), 7.49 (t, *J* = 7.6 Hz, 1H, Ar-H), 7.45 (d, *J* = 7.6 Hz, 1H, Ar-H), 4.90 (d, *J* = 1.8 Hz, 1H, H-1), 4.86, 4.63 (A, B of AB, *J* = 12.0 Hz, 2H, CH_2Ar), 3.94 (s, 3H, OMe), 3.89–3.87 (m, 2H, H-2, H-3), 3.79–3.73 (m, 2H, H-4, H-6a), 3.68–3.64 ppm (m, 2H, H-5, H-6b); ^{13}C NMR (125 MHz, CD_3OD): δ = 168.42 (CO), 146.91, 141.31, 139.97, 131.13, 130.20, 129.07, 128.17, 127.91, 127.67 (12C, Ar-C), 100.76 (C-1), 75.02 (C-5), 72.65 (C-3), 72.22 (C-2), 69.73 (CH_2Ar), 68.65 (C-4), 62.98 (C-6), 52.66 ppm (OMe); HRMS: *m/z*: calcd for $C_{21}H_{24}NaO_8$ $[M+Na]^+$: 427.1363, found: 427.1361.

Methyl 4′-chloro-3′-[(α-D-mannopyranosyloxy)methyl]biphenyl-4-carboxylate (19b): Prepared according to general procedure C from **18b** (40 mg, 0.066 mmol). Yield: 26 mg (90%). R_f = 0.19 (CH_2Cl_2/MeOH, 8:1); $[\alpha]_D^{20}$ +101.8 (*c* = 0.50, MeOH); ^{1}H NMR (500 MHz, CD_3OD): δ = 8.06 (d, *J* = 8.4 Hz, 2H, Ar-H), 7.69 (d, *J* = 8.4 Hz, 2H, Ar-H), 7.57–7.56 (m, 2H, Ar-H), 7.41 (s, 1H, Ar-H), 4.87 (s, 1H, H-1), 4.80, 4.58 (A, B of AB, *J* = 12.3 Hz, 2H, CH_2Ar), 3.91 (s, 3H, OMe), 3.87–3.83 (m, 2H, H-2, H-3), 3.74–3.57 ppm (m, 4H, H-4, H-5, H-6); ^{13}C NMR (125 MHz, CD_3OD): δ = 168.74 (CO), 145.78, 143.68, 142.71, 136.55, 131.75, 131.31, 129.02, 128.77, 127.95, 126.63 (12C, Ar-C), 101.47 (C-1), 75.65 (C-5), 73.16 (C-3), 72.65 (C-2), 69.49 (CH_2Ar), 69.13 (C-4), 63.49 (C-6), 53.26 ppm (OMe); HRMS: *m/z*: calcd for $C_{21}H_{23}ClNaO_8$ $[M+Na]^+$: 461.0974, found: 461.0975.

Methyl 2′-[(α-D-mannopyranosyloxy)methyl]biphenyl-4-carboxylate (22): Prepared according to general procedure C from **21** (48 mg, 0.084 mmol). Yield: 16 mg (47%). R_f = 0.42 (CH_2Cl_2/MeOH, 8:1); $[\alpha]_D^{20}$ +61.9 (*c* = 0.90, MeOH); ^{1}H NMR (500 MHz, CD_3OD): δ = 8.11–8.09 (m, 2H, Ar-H), 7.57 (m, 1H, Ar-H), 7.51–7.49 (m, 2H, Ar-H), 7.43–7.40 (m, 2H, Ar-H), 7.31 (m, 1H, Ar-H), 4.71 (A of AB, *J* = 11.4 Hz, 1H, CH_2Ar), 4.70 (d, *J* = 1.5 Hz, 1H, H-1), 4.38 (B of AB, *J* = 11.4 Hz, 1H, CH_2Ar), 3.75–3.60 (m, 5H, H-2, H-3, H-4, H-6), 3.95 (s, 3H, OMe), 3.40 ppm (ddd, *J* = 3.0, 5.6, 6.8 Hz, 1H, H-5); ^{13}C NMR

(125 MHz, CD_3OD): δ = 168.42 (CO), 147.33, 142.66, 136.03, 130.83, 130.53, 130.47, 129.26, 129.15 (12C, Ar-C), 101.14 (C-1), 74.78 (C-5), 72.60 (C-3), 72.18 (C-2), 68.37 (2C, C-4, CH_2Ar), 62.71 ppm (C-6), 52.69 (OMe); HRMS: *m/z*: calcd for $C_{21}H_{24}NaO_8Na$ $[M+Na]^+$: 427.1363, found: 427.1367.

Sodium 3′-[(α-D-mannopyranosyloxy)methyl]biphenyl-4-carboxylate (6): Prepared according to general procedure D from **18a** (35 mg, 0.061 mmol). Yield: 24 mg (96%). $[\alpha]_D^{20}$ +64.5 (*c* = 0.30, MeOH/H_2O 1:1); ^{1}H NMR (500 MHz, D_2O): δ = 7.80–7.78 (m, 2H, Ar-H), 7.50–7.43 (m, 4H, Ar-H), 7.31–7.24 (m, 2H, Ar-H), 4.82 (s, 1H, H-1), 4.58, 4.40 (A, B of AB, *J* = 11.5 Hz, 2H, CH_2Ar), 3.82 (m, 1H, H-2), 3.75–3.50 ppm (m, 5H, H-3, H-4, H-5, H-6); ^{13}C NMR (125 MHz, D_2O): δ = 175.14 (CO), 142.69, 140.05, 137.34, 135.01, 129.46, 129.28, 127.92, 126.87, 126.64 (12C, Ar-C), 99.40 (C-1), 72.84 (C-5), 70.51 (C-3), 70.01 (C-2), 69.29 (CH_2Ar), 66.61 (C-4), 60.71 ppm (C-6); HRMS: *m/z*: calcd for $C_{20}H_{22}NaO_8$ $[M+Na]^+$: 413.1207, found: 413.1211.

Sodium 4′-chloro-3′-[(α-D-mannopyranosyloxy)methyl]biphenyl-4-carboxylate (20): Prepared according to general procedure D from **18b** (54 mg, 0.089 mmol). Yield: 4 mg (10%). $[\alpha]_D^{20}$ +44.7 (*c* = 0.30, MeOH); ^{1}H NMR (500 MHz, D_2O): δ = 7.86 (d, *J* = 7.8 Hz, 2H, Ar-H), 7.58–7.56 (m, 3H, Ar-H), 7.46, 7.34 (2 s, 2H, Ar-H), 4.90 (s, 1H, H-1), 4.58, 4.50 (A, B of AB, *J* = 12.3 Hz, 2H, CH_2Ar), 3.91 (s, 1H, H-2), 3.78–3.75 (m, 2H, H-3, H-4), 3.71–3.59 ppm (m, 3H, H-5, H-6); ^{13}C NMR (125 MHz, D_2O): δ = 174.76 (CO), 141.82, 141.55, 139.40, 134.37, 129.56, 127.34, 126.74, 126.62, 125.15 (12C, Ar-C), 99.99 (C-1), 72.96 (C-5), 70.55 (C-3), 70.04 (C-2), 68.72 (CH_2Ar), 66.66 (C-4), 60.77 ppm (C-6); HRMS: *m/z*: calcd for $C_{20}H_{21}ClNaO_8$ $[M+Na]^+$: 447.0817, found: 447.0816.

Sodium 2′-[(α-D-mannopyranosyloxy)methyl]biphenyl-4-carboxylate (23): Prepared according to general procedure D from **21** (78 mg, 0.137 mmol). Yield: 26 mg (46%). $[\alpha]_D^{20}$ +53.2 (*c* = 0.40, MeOH); ^{1}H NMR (500 MHz, D_2O): δ = 7.91–7.89 (m, 2H, Ar-H), 7.43–7.34 (m, 5H, Ar-H), 7.26 (m, 1H, Ar-H), 4.68 (s, 1H, H-1), 4.57, 4.31 (A, B of AB, *J* = 10.8 Hz, 2H, CH_2Ar), 3.57 (m, 1H, H-2), 3.46–3.39 (m, 4H, H-3, H-4, H-6), 2.83 ppm (m, 1H, H-5); ^{13}C NMR (125 MHz, D_2O): δ = 173.20 (CO), 144.48, 141.80, 133.47, 132.43, 130.69, 129.95, 129.27, 128.96, 128.32 (12C, Ar-C), 99.90 (C-1), 72.44 (C-5), 70.33 (C-3), 69.82 (C-2), 68.14 (CH_2Ar), 65.99 (C-4), 60.25 ppm (C-6); HRMS: *m/z*: calcd for $C_{20}H_{22}NaO_8$ $[M+Na]^+$: 413.1207, found: 413.1208.

4-(4,4,5,5-Tetramethyl)-1,3,2-dioxaborolan-2-yl)phenyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (27): A microwave tube was charged with **26**[37] (240 mg, 0.55 mmol), KOAc (161 mg, 1.65 mmol), bis(pinacolato)diborone (152 mg, 0.60 mmol) and Pd(dppf)Cl_2·CH_2Cl_2 (13 mg, 0.017 mmol). The tube was closed, evacuated and flushed with argon. Then anhydrous DMF (1 mL) was added under a stream of argon. The mixture was degassed in an ultrasonic bath and flushed with argon for 5 min, and then heated by microwave irradiation at 120 °C for 2 h. The reaction mixture was cooled to RT and diluted with CH_2Cl_2/H_2O (100 mL, 1:1). The organic layer was washed with H_2O (50 mL) and brine (50 mL), dried over Na_2SO_4 and concentrated. The residue was purified by MPLC (toluene/EtOAc, 4:1) to afford **27** (120 mg, 50%) as colorless oil. $[\alpha]_D^{20}$ +58.1 (*c* = 0.60, EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.76 (d, *J* = 8.6 Hz, 2H, Ar-H), 7.08 (d, *J* = 8.6 Hz, 2H, Ar-H), 5.58–5.55 (m, 2H, H-1, H-3), 5.45 (dd, *J* = 1.9, 3.4 Hz, 1H, H-2), 5.37 (t, *J* = 10.0 Hz, 1H, H-4), 4.28 (dd, *J* = 5.0, 12.0 Hz, 1H. H-6a), 4.05–4.02 (m, 2H, H-6b, H-5), 2.20, 2.05, 2.03 (3 s, 12H, 4 OAc), 1.33 ppm (s, 12H, 4 CH_3); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 170.55, 169.91, 169.74 (4C, 4 CO), 157.98, 136.62, 136.58, 115.67 (5C, Ar-C), 95.44 (C-1), 83.77 (Ar-C), 69.37 (C-2), 69.21 (C-5), 68.87 (C-3), 65.92 (C-4), 62.06 (C-6), 24.86, 24.58 (4C, 4 CH_3), 20.87, 20.69 ppm (4C, 4 $COCH_3$); ESI-MS: *m/z*: calcd for $C_{26}H_{35}BNaO_{12}$ $[M+Na]^+$: 573.21, found: 573.32.

Methyl 2-[4′-(2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyloxy)biphenyl-4-yl]acetate (29): Prepared according to general procedure B from methyl 2-(4-bromophenyl)acetate (**28**, 41.2 mg, 0.180 mmol), **27** (109 mg, 0.198 mmol), Pd(dppf)Cl_2·CH_2Cl_2 (4.4 mg, 5.4 μmol) and K_3PO_4 (57.3 mg, 0.270 mmol). Yield: 35 mg (34%) as yellow oil. R_f = 0.25 (petrol ether/EtOAc 2:1); $[\alpha]_D^{20}$ +75.09 (*c* = 0.8, EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.52–7.49 (m, 4H, Ar-H), 7.35–7.33 (m, 2H, Ar-H), 7.17–7.14 (m, 2H, Ar-H), 5.60–5.56 (m, 2H, H-1, H-3), 5.47 (dd, *J* = 1.8, 3.5 Hz, 1H, H-2), 5.38 (t, *J* = 10.0 Hz, 1H, H-4), 4.29 (dd, *J* = 5.0, 11.9 Hz, 1H, H-6a), 4.15–4.08 (m, 2H, H-6b, H-5), 3.71 (s, 3H, OMe), 3.66 (s, 2H, $ArCH_2$), 2.21, 2.06, 2.05, 2.03 ppm (4 s, 12H, 4 OAc); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 171.99, 170.53, 169.99, 169.95, 169.76 (5 CO), 155.09, 139.26, 135.72, 132.83, 129.73, 128.21, 127.03, 116.82 (12C, Ar-C), 95.87 (C-1), 69.43 (C-2), 69.23 (C-5), 68.91 (C-3), 65.99 (C-4), 62.15 (C-6), 52.11 (OMe), 40.78 ($ArCH_2$), 20.88, 20.71, 20.70, 20.67 ppm (4 $COCH_3$); ESI-MS: *m/z*: calcd for $C_{29}H_{32}NaO_{12}$ $[M+Na]^+$: 595.18, found: 595.21.

Methyl 2-[4′-(2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyloxy)biphenyl-4-yl]cyclopropanecarboxylate (33): Prepared according to general procedure B from methyl 1-(4-bromophenyl)cyclopropanecarboxylate (**32**, 42.6 mg, 0.167 mmol), **27** (101 mg, 0.184 mmol), Pd(dppf)Cl_2·CH_2Cl_2 (4.1 mg, 5.0 μmol) and K_3PO_4 (53.2 mg, 0.251 mmol). Yield: 60 mg (56%) as colorless oil. R_f = 0.31 (PE/EtOAc, 2:1); $[\alpha]_D^{20}$ +70.2 (*c* = 1.00, EtOAc); ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.54–7.48 (m, 4H, Ar-H), 7.40–7.39 (m, 2H, Ar-H), 7.17–7.14 (m, 2H, Ar-H), 5.59 (dd, *J* = 3.55, 10.1 Hz, 1H, H-3), 5.56 (d, *J* = 1.6 Hz, 1H, H-1), 5.46 (dd, *J* = 1.9, 3.5 Hz, 1H, H-2), 5.38 (t, *J* = 10.0 Hz, 1H, H-4), 4.29 (dd, *J* = 5.1, 12.0 Hz, 1H, H-6a), 4.15–4.09 (m, 2H, H-6b, H-5), 3.65 (s, 3H, OMe), 2.21, 2.06, 2.05, 2.03 (4 s, 12H, 4 OAc), 1.64–1.62 (m, 2H, *c*Pr), 1.27–1.16 ppm (m, 2H, *c*Pr); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 175.04, 170.53, 169.98, 169.95, 169.75 (5 CO), 155.10, 139.25, 138.43, 135.76, 130.94, 128.24, 126.61, 116.80 (12C, Ar-C), 95.89 (C-1), 69.44 (C-5), 69.23 (C-2), 68.90 (C-3), 66.00 (C-4), 62.15 (C-6), 52.42 (OMe), 28.67 (*c*Pr), 20.71, 20.68 (4C, 4 $COCH_3$), 16.75 ppm (*c*Pr); ESI-MS: *m/z*: calcd for $C_{31}H_{34}NaO_{12}$ $[M+Na]^+$: 621.19, found: 621.26.

Methyl 2-[4′-(α-D-mannopyranosyloxy)biphenyl-4-yl]acetate (30): Prepared according to general procedure C from **29** (30 mg, 0.052 mmol). Yield: 20 mg (95%). R_f = 0.25 (CH_2Cl_2/MeOH, 8:1); $[\alpha]_D^{20}$ +116.0 (*c* = 0.50, MeOH); ^{1}H NMR (500 MHz, CD_3OD): δ = 7.57–7.53 (m, 4H, Ar-H), 7.34–7.33 (m, 2H, Ar-H), 7.22–7.20 (m, 2H, Ar-H), 5.54 (d, *J* = 1.5 Hz, 1H, H-1), 4.05 (dd, *J* = 1.8, 3.3 Hz, 1H, H-2), 3.95 (dd, *J* = 3.4, 9.5 Hz, 1H, H-3), 3.82–3.74 (m, 3H, H-4, H-6), 3.71 (s, 3H, OMe), 3.66 (s, 2H, $ArCH_2$), 3.65 ppm (ddd, *J* = 2.5, 5.2, 9.7 Hz, 1H, H-5); ^{13}C NMR (125 MHz, CD_3OD): δ = 174.02 (CO), 157.50, 140.77, 136.22, 134.29, 130.81, 129.00, 127.77, 118.13 (12C, Ar-C), 100.23 (C-1), 75.42 (C-5), 72.45 (C-3), 72.03 (C-2), 68.38 (C-4), 62.70 (C-6), 52.49 (OMe), 41.34 ppm ($ArCH_2$); HRMS: *m/z*: calcd for $C_{21}H_{24}NaO_8$ $[M+Na]^+$: 427.1363, found: 427.1363.

Methyl 2-[4′-(α-D-mannopyranosyloxy)biphenyl-4-yl]cyclopropanecarboxylate (34): Prepared according to general procedure C from **33** (38 mg, 0.063 mmol). Yield: 9 mg (33%). R_f = 0.33 (CH_2Cl_2/MeOH, 8:1); $[\alpha]_D^{20}$ +108.0 (*c* = 0.30, MeOH); ^{1}H NMR (500 MHz, CD_3OD): δ = 7.46–7.39 (m, 4H, Ar-H), 7.28–7.26 (m, 2H, Ar-H), 7.10–7.07 (m, 2H, Ar-H), 5.42 (d, *J* = 1.7 Hz, 1H, H-1), 3.93 (dd, *J* = 1.9, 3.4 Hz, 1H, H-2), 3.82 (dd, *J* = 3.4, 9.4 Hz, 1H, H-3), 3.69–3.61 (m, 3H, H-4, H-6), 3.53 (m, 4H, OMe, H-5), 1.49–1.47 (m, 2H, *c*Pr), 1.14–1.11 ppm (m, 2H, *c*Pr); ^{13}C NMR (125 MHz, CD_3OD): δ = 157.50, 140.87, 139.51, 136.26, 132.03, 129.04, 127.43, 118.11 (12C, Ar-C),

100.20 (C-1), 75.43 (C-5), 72.42 (C-3), 72.02 (C-2), 68.34 (C-4), 62.68 (C-6), 52.81 (OMe), 17.20 ppm (2C, *c*Pr); HRMS: *m/z*: calcd for $C_{23}H_{26}NaO_8$ $[M+Na]^+$: 453.1520, found: 453.1523.

Sodium 2-[4′-(α-D-mannopyranosyloxy)biphenyl-4-yl]acetate (31): Prepared according to general procedure D from **29** (59 mg, 0.103 mmol). Yield: 17 mg (40%). $[\alpha]_D^{20}$ +94.0 (c=0.20, MeOH/H_2O 1:1); ^{1}H NMR (500 MHz, D_2O): δ=7.61 (d, J=8.6 Hz, 2H, Ar-H), 7.55 (d, J=8.0 Hz, 2H, Ar-H), 7.31 (d, J=8.0 Hz, 2H, Ar-H), 7.19 (d, J=8.6 Hz, 2H, Ar-H), 5.60 (s, 1H, H-1), 4.13 (m, 1H, H-2), 4.00 (dd, J=3.2, 8.5 Hz, 1H, H-3), 3.75–3.67 (m, 4H, H-4, H-5, H-6), 3.51 ppm (s, 2H, $ArCH_2$); ^{13}C NMR (125 MHz, D_2O): δ=154.94, 137.93, 136.29, 135.08, 129.76, 128.07, 126.72, 117.49 (12C, Ar-C), 98.20 (C-1), 73.37 (C-5), 70.40 (C-3), 69.89 (C-2), 66.58 (C-4), 60.65 (C-6), 43.89 ppm ($ArCH_2$); HRMS: *m/z*: calcd for $C_{20}H_{22}NaO_8$ $[M+Na]^+$: 413.1207, found: 413.1208.

Sodium 2-[4′-(α-D-mannopyranosyloxy)biphenyl-4-yl]cyclopropanecarboxylate (35): Prepared according to general procedure D from **33** (59 mg, 0.099 mmol). Yield: 10 mg (23%). $[\alpha]_D^{20}$ +95.0 (c=0.20, dioxane/H_2O 1:1); ^{1}H NMR (500 MHz, D_2O): δ=7.62–7.60 (m, 2H, Ar-H), 7.54–7.53 (m, 2H, Ar-H), 7.38–7.19 (m, 4H, Ar-H), 5.60 (s, 1H, H-1), 4.13 (m, 1H, H-2), 4.00 (m, 1H, H-3), 3.75–3.67 (4H, H-4, H-5, H-6), 1.33 (s, 2H, *c*Pr), 1.01 ppm (s, 2H, *c*Pr); ^{13}C NMR (125 MHz, D_2O): δ=128.67, 126.10, 124.37, 115.47 (12C, Ar-C), 96.18 (C-1), 71.35 (C-5), 68.38 (C-3), 67.87 (C-2), 64.56 (C-4), 58.62 (C-6), 12.66 ppm (2C, *c*Pr); HRMS: *m/z*: calcd for $C_{22}H_{24}NaO_8$ $[M+Na]^+$: 439.1363, found: 439.1363.

Competitive binding assay

A recombinant protein consisting of the CRD of FimH linked with a thrombin cleavage site (Th) to a His_6-tag (FimH-CRD-Th-His_6) was expressed in *E. coli* strain HM125 and purified by affinity chromatography.[16] To determine the affinity of the various FimH antagonists, a competitive binding assay described previously[16] was applied. Microtiter plates (F96 MaxiSorp, Nunc) were coated with a 10 μg mL^{-1} solution of FimH-CRD-Th-His_6 in 20 mM HEPES, 150 mM NaCl, and 1 mM $CaCl_2$, pH 7.4 (assay buffer), 100 μL per well, overnight at 4 °C. The coating solution was discarded, and the wells were blocked with 3% BSA in assay buffer (150 μL per well) for 2 h at 4 °C. After three washing steps with assay buffer (150 μL per well), a fourfold serial dilution of the test compound (50 μL per well) in assay buffer containing 5% DMSO and streptavidin-peroxidase coupled Man-α(1–3)[Man-α(1–6)]-Man-β(1–4)-Glc*N*Ac-β(1–4)-Glc*N*Acβ polyacrylamide (TM-PAA) polymer (50 μL per well of a 0.5 μg mL^{-1} solution) were added. On each individual microtiter plate, *n*-heptyl α-D-mannopyranoside (**1**) was tested in parallel. The plates were incubated for 3 h at 25 °C and 350 rpm and then carefully washed four times with 150 μL per well assay buffer. After the addition of 100 μL per well of 2,2′-azino-di-(3-ethylbenzthiazoline-6-sulfonic acid) (ABTS) substrate, the colorimetric reaction was allowed to develop for 4 min and then was stopped by the addition of 2% aqueous oxalic acid before the optical density (OD) was measured at 415 nm on a microplate reader (Spectramax 190, Molecular Devices, CA, USA). The IC_{50} values of the compounds tested in duplicate were calculated with Prism software (GraphPad Software Inc., La Jolla, CA, USA). The IC_{50} defines the molar concentration of the test compound that decreases the maximal specific binding of TM-PAA polymer to FimH-CRD by 50%. The relative IC_{50} (rIC_{50}) is the ratio of the IC_{50} of the test compound to the IC_{50} of *n*-heptyl α-D-mannopyranoside (**1**).

Cell-based flow cytometry assay

The assay was performed as described previously.[17] Briefly, 5637 cells (DSMZ, Braunschweig, Germany) were grown to confluence in 24-well plates. Before infection, a serial dilution of test compound in 5% DMSO, PBS (Sigma–Aldrich) was prepared. GFP-labeled UTI89 bacteria (200 μL) in RPMI 1640 medium (Invitrogen, Basel, Switzerland) were pre-incubated with test compound (25 μL) for 10 min at RT. The bacteria–antagonist mixtures were then added to the monolayers of 5637 cells. The multiplicity of infection (MOI) was 1:50 (cell/bacteria). To homogenize the infection, plates were centrifuged at RT for 3 min at 600 *g*. After an incubation time of 1.5 h at 37 °C, infected cells were washed four times with RPMI 1640 medium and suspended in ice-cold PBS for 5–20 min (treatment with ice-cold PBS results in the detachment of the infected cells). Cells were then kept in the dark until analysis. Samples were measured with a CyAn ADP flow cytometer (Beckman–Coulter, Brea, CA, USA) and analyzed by gating on the eukaryotic cells based on forward (FSC) and side scatter (SSC), which excludes unbound labeled bacteria and debris from analysis. A total of 10^4 cells were measured per sample. Data were acquired in a linear mode for the SSC and logarithmic mode for FSC and the green fluorescent channel FL1-H (GFP). The mean fluorescence intensity (MFI) of GFP was counted as a surrogate marker for the adherence of bacteria. Quantification of adhesion was evaluated with the FlowJo software 9.0.1 (Tree Star Inc., Ashland, OR, USA). IC_{50} values were determined by plotting the concentration of the antagonist in a logarithmic mode versus the MFI and by fitting the curve with Prism software (GraphPad, inhibition curve, variable slope), (n=2–3, in duplicate/triplicate).

Isothermal titration calorimetry (ITC)

For the ITC experiments, the His tag in FimH-CRD-Th-His_6 was cleaved.[16] Briefly, the protein (1 mg) was incubated with 10 U thrombin (T-6884, Sigma–Aldrich) in 20 mM Tris·HCl, pH 8.4, 150 mM NaCl and 2.5 mM $CaCl_2$ (cleavage buffer) at 20 °C for 16 h. The mixture was then applied to a gel filtration column (Bio-Prep SE100/17, Bio-Rad) attached to an FPLC system. The chromatography was run with assay buffer and analyzed by SDS-PAGE. The fractions containing FimH-CRD were pooled and concentrated by ultrafiltration (MWCO10, Sartorius AG, Tagelswangen, Switzerland).

The ITC experiments were performed using a VP-ITC instrument from MicroCal Inc. (GE Healthcare, Northampton, MA, USA). The measurements were performed at 25 °C. Prior to measurements, the protein was dialyzed in assay buffer (10 mM HEPES, 150 mM NaCl, 1 mM $CaCl_2$, pH 7.4 (HBS-Ca). Injections of 3–5 μL ligand solutions (150 μM) were added at an interval of 10 min into the sample cell solution containing FimH-CRD (8–22 μM, sample cell volume 1.4523 mL) with stirring at 307 rpm. Protein concentration was determined by HPLC-UV against a BSA standard.[38] The quantity c=$Mt(0)$ K_D^{-1}, where $Mt(0)$ is the initial macromolecule concentration, is of importance in titration microcalorimetry. The c values ranged between 300 and 3200. Because the smallest reliable volumes were injected, sigmoidal curves were obtained. Control experiments injecting ligand solution into buffer without protein showed that the heat of dilution was small and constant. Baseline correction and peak integration were accomplished using Origin 7 as described by the manufacturer (OriginLab, Northampton, MA, USA). The first injection was always excluded from data analysis because it usually suffers from sample loss during the mounting of the syringe and the equilibration preceding the actual titration. A three-parameter (N (stoichiometry), K_D (dissociation constant) and ΔH° (change in enthalpy) nonlinear least-square data fitting was per-

formed in a Microsoft Excel spreadsheet using the Solver add-in (Frontline System)[39,40] according to binding isotherms published by Ziegler and Seelig.[41]

Thermodynamics parameters were calculated from Equation (4).

$$\Delta G = \Delta H - T\Delta S = RT \ln K_D = -RT \ln K_A \quad (4)$$

where ΔG, ΔH, and ΔS are the changes in free energy, enthalpy, and entropy of binding, respectively, T is the absolute temperature, and R is the universal gas constant (8.314 $J\,mol^{-1}\,K^{-1}$).

Determination of pharmacokinetic parameters

Materials: Dimethyl sulfoxide (DMSO), 1-octanol, Dulbecco's modified Eagle's medium (DMEM) high glucose, L-glutamine solution, penicillin–streptomycin solution, Dulbecco's phosphate-buffered saline (DPBS), and trypsin–EDTA solution were purchased from Sigma–Aldrich. MEM nonessential amino acid (MEM-NEAA) solution, fetal bovine serum (FBS), and DMEM without sodium pyruvate and phenol red were bought from Invitrogen. PAMPA System Solution, GIT-0 Lipid Solution, and Acceptor Sink Buffer were ordered from plon (Woburn, MA, USA). Acetonitrile (MeCN) was bought from Acros Organics. The Caco-2 cells were kindly provided by Prof. G. Imanidis, FHNW, Muttenz, Switzerland and originated from the American Type Culture Collection (Rockville, MD, USA).

Parallel artificial membrane permeation assay (PAMPA)

Values of $\log P_e$ were determined in a 96-well format with the PAMPA[33] permeation assay. For each compound, measurements were performed at three pH values (5.0, 6.2, 7.4) in quadruplicate. For this purpose, 12 wells of a deep-well plate, i.e., four wells per pH value, were filled with 650 μL System Solution. Samples (150 μL) were withdrawn from each well to determine the blank spectra by UV spectroscopy (SpectraMax 190). Then, analyte dissolved in DMSO was added to the remaining System Solution to yield 50 μM solutions. To exclude precipitation, the optical density was measured at 650 nm, with 0.01 being the threshold value. Solutions exceeding this threshold were filtered. Afterward, samples (150 μL) were withdrawn to determine the reference spectra. Further 200 μL were transferred to each well of the donor plate of the PAMPA sandwich P/N 110 163 (plon, Woburn MA, USA). The filter membranes at the bottom of the acceptor plate were impregnated with 5 μL of GIT-0 Lipid Solution, and 200 μL of Acceptor Sink Buffer were filled into each acceptor well. The sandwich was assembled, placed in the GutBox, and left undisturbed for 16 h. It was then disassembled, and samples (150 μL) were transferred from each donor and acceptor well to UV plates. Quantification was performed by both UV spectroscopy and LC–MS; $\log P_e$ values were calculated with the aid of the PAMPA Explorer Software (plon, version 3.5).

Colorectal adenocarcinoma (Caco-2) cell permeation assay

Caco-2 cells were cultivated in tissue culture flasks (BD Biosciences, Franklin Lakes, NJ, USA) with DMEM high-glucose medium containing L-glutamine (2 mM), nonessential amino acids (0.1 mM), penicillin (100 $U\,mL^{-1}$), streptomycin (100 $\mu g\,mL^{-1}$), and FBS (10%). The cells were kept at 37 °C in humidified air containing 5% CO_2, and the medium was changed every second day. When ~90% confluence was reached, the cells were split in a 1:10 ratio and distributed to new tissue culture flasks. At passage numbers between 60 and 65, they were seeded at a density of 5.3×10^5 cells per well to Transwell 6-well plates (Corning Inc., Corning, NY, USA) with 2.5 mL culture medium in the basolateral and 1.8 mL in the apical compartment. The medium was renewed on alternate days. Permeation experiments were performed between days 19 and 21 post-seeding. Prior to the experiment, the integrity of the Caco-2 monolayers was evaluated by measuring the transepithelial electrical resistance (TEER) with an Endohm tissue resistance instrument (World Precision Instruments Inc., Sarasota, FL, USA). Only wells with TEER values $>300\ \Omega c\,m^2$ were used. Experiments were performed in the apical-to-basolateral (absorptive) and basolateral-to-apical (secretory) directions in triplicate. Transport medium (DMEM without sodium pyruvate and phenol red) was withdrawn from the donor compartments of three wells and replaced by the same volume of compound stock solutions to reach an initial sample concentration of 62.5 μM. The Transwell plate was then shaken (250 rpm) in the incubator. Samples (40 μL) were withdrawn after 15, 30, and 60 min from the donor and acceptor compartments, and their concentrations were determined by LC–MS. Apparent permeability coefficients (P_{app}) were calculated according to the equation

$$P_{app} = \frac{dQ}{dt} \times \frac{1}{A \times c_0} \quad (5)$$

where dQ/dt is the permeability rate, A the surface area of the monolayer, and c_0 the initial concentration in the donor compartment.[42] After the experiment, TEER values were assessed again for each well and results from wells with values $<300\ \Omega c\,m^2$ were discarded.

log $D_{7.4}$ determination

The in silico prediction tool ALOGPS[43] was used to estimate the $\log P$ values of the compounds. Depending on these values, the compounds were classified into three categories: hydrophilic compounds ($\log P < 0$), moderately lipophilic compounds ($0 \le \log P \le 1$) and lipophilic compounds ($\log P > 1$). For each category, two different ratios (volume of 1-octanol to volume of buffer) were defined as experimental parameters (Table 4).

Table 4. Compound classification based on estimated $\log P$ values.[43]

Compound type	$\log P$	Ratio (1-octanol)/(buffer)
hydrophilic	<0	30:140, 40:130
moderately lipophilic	0–1	70:110, 110:70
lipophilic	>1	3:180, 4:180

Equal amounts of phosphate buffer (0.1 M, pH 7.4) and 1-octanol were mixed and shaken vigorously for 5 min to saturate the phases. The mixture was left until separation of the two phases occurred, and the buffer was retrieved. Stock solutions of the test compounds were diluted with buffer to a concentration of 1 μM. For each compound, six determinations, i.e., three determinations per 1-octanol/buffer ratio, were performed in different wells of a 96-well plate. The respective volumes of buffer containing analyte (1 μM) were pipetted to the wells and covered by saturated 1-octanol according to the chosen volume ratio. The plate was sealed with aluminum foil, shaken (1350 rpm, 25 °C, 2 h) on a Heidoph Titramax 1000 plate shaker (Heidolph Instruments GmbH & Co. KG, Schwabach, Germany) and centrifuged (2000 rpm, 25 °C,

5 min, 5804 R Eppendorf centrifuge, Hamburg, Germany). The aqueous phase was transferred to a 96-well plate for analysis by LC–MS.

$\log D_{7.4}$ was calculated from the 1-octanol/buffer ratio (o/b), the initial concentration of the analyte in buffer (1 μM), and the concentration of the analyte in the aqueous phase (c_B) with equation:

$$\log D_{7.4} = \log\left(\frac{1\mu M - c_B}{c_B} \times \frac{1}{o:b}\right) \quad (6)$$

Solubility

Solubility was determined in a 96-well format using the μSOL Explorer solubility analyzer (pIon, version 3.4.0.5). For each compound, measurements were performed at three pH values (3.0, 5.0, 7.4) in triplicates. For this purpose, nine wells of a deep-well plate, that is, three wells per pH value, were filled with 300 μL of an aqueous universal buffer solution. Aliquots (3 μL) of a compound stock solution (10–50 mM in DMSO) were added and thoroughly mixed. The final sample concentration was 0.1–0.5 mM, the residual DMSO concentration was 1.0% (*v*/*v*) in the buffer solutions. After 15 h, the solutions were filtered (0.2 μm 96-well filter plates) using a vacuum to collect manifold (Whatman Ltd., Maidstone, UK) to remove any precipitates. Equal amounts of filtrate and *n*-propanol were mixed and transferred to a 96-well plate for UV detection (190–500 nm). The amount of material dissolved was calculated by comparison with UV spectra obtained from reference samples, which were prepared by dissolving compound stock solution in a 1:1 mixture of buffer and *n*-propanol (final concentrations 0.017–0.083 mM).

LC–MS measurements

Analyses were performed using an 1100/1200 Series HPLC System coupled to a 6410 Triple Quadrupole mass detector (Agilent Technologies, Inc., Santa Clara, CA, USA) equipped with electrospray ionization. The system was controlled with the Agilent MassHunter Workstation Data Acquisition software (version B.01.04). The column used was an Atlantis T3 C_{18} column (2.1×50 mm) with a 3 μm particle size (Waters Corp., Milford, MA, USA). The mobile phase consisted of two eluents: solvent A (H_2O, containing 0.1% formic acid, *v*/*v*) and solvent B (MeCN, containing 0.1% formic acid, *v*/*v*), both delivered at 0.6 mL min^{-1}. The gradient was ramped from 95% A/5% B to 5% A/95% B over 1 min, and then held at 5% A/95% B for 0.1 min. The system was then brought back to 95% A/5% B, resulting in a total duration of 4 min. MS parameters such as fragmentor voltage, collision energy, and polarity were optimized individually for each analyte, and the molecular ion was followed for each compound in the multiple reaction monitoring mode. The concentrations of the analytes were quantified by the Agilent Mass Hunter Quantitative Analysis software (version B.01.04).

Abbreviations

Caco-2 cells, colorectal adenocarcinoma cells; CRD, carbohydrate recognition domain; *D*, distribution coefficient octanol/H_2O; GFP, green fluorescent protein; HPLC, high-performance liquid chromatography; IC_{50}, half-maximal inhibitory concentration; ITC, isothermal titration calorimetry; MFI, mean fluorescence intensity; PAMPA, parallel artificial membrane permeability assay; P_{app}, apparent permeability coefficient; P_e, effective permeation value; SAR, structure–activity relationship; SPR, structure–property relationship; UPEC, uropathogenic *E. coli*; UTI, urinary tract infection.

Acknowledgement

Financial support from the Swiss National Science Foundation (SNF interdisciplinary grant K-32K1-120904) is gratefully acknowledged.

Keywords: bacterial adhesin · FimH antagonists · flow cytometry · isothermal titration calorimetry · urinary tract infections

[1] a) T. M. Hooton, W. E. Stamm, *Infect. Dis. Clin. North Am.* **1997**, *11*, 551–581; b) T. J. Wiles, R. R. Kulesus, M. A. Mulvey, *Exp. Mol. Pathol.* **2008**, *85*, 11–19; c) S. D. Fihn, *N. Engl. J. Med.* **2003**, *349*, 259–266.
[2] C. Svanborg, G. Godaly, *Infect. Dis. Clin. North Am.* **1997**, *11*, 513–529.
[3] a) J. D. Schilling, S. J. Hultgren, *Int. J. Antimicro. Ag.* **2002**, *19*, 457–460; b) M. G. Blango, M. A. Mulvey, *Antimicrob. Agents Chemother.* **2010**, *54*, 1855–1863.
[4] G. Capitani, O. Eidam, R. Glockshuber, M. G. Grutter, *Microbes Infect.* **2006**, *8*, 2284–2290.
[5] a) M. A. Mulvey, *Cell Microbiol.* **2002**, *4*, 257–271; b) S. G. Gouin, A. Wellens, J. Bouckaert, J. Kovensky, *ChemMedChem* **2009**, *4*, 749–755.
[6] N. Sharon, *Biochim. Biophys. Acta.* **2006**, *1760*, 527–537.
[7] a) N. Firon, I. Ofek, N. Sharon, *Biochem. Biophys. Res. Commun.* **1982**, *105*, 1426–1432; b) N. Firon, I. Ofek, N. Sharon, *Carbohydr. Res.* **1983**, *120*, 235–249; c) N. Sharon, *FEBS Lett.* **1987**, *217*, 145–157.
[8] D. Choudhury, A. Thompson, V. Stojanoff, S. Langermann, J. Pinkner, S. J. Hultgren, S. D. Knight, *Science* **1999**, *285*, 1061–1066.
[9] a) C. S. Hung, J. Bouckaert, D. Hung, J. Pinkner, C. Widberg, A. DeFusco, C. G. Auguste, R. Strouse, S. Langermann, G. Waksman, S. J. Hultgren, *Mol. Microbiol.* **2002**, *44*, 903–915; b) J. Bouckaert, J. Berglund, M. Schembri, E. D. Genst, L. Cools, M. Wuhrer, C. S. Hung, J. Pinkner, R. Slättergard, A. Zavialov, D. Choudhury, S. Langermann, S. J. Hultgren, L. Wyns, P. Klemm, S. Oscarson, S. D. Knight, H. De Greve, *Mol. Microbiol.* **2005**, *55*, 441–455; c) A. Wellens, C. Garofalo, H. Nguyen, N. Van Gerven, R. Slättergard, J.-P. Hernalsteens, L. Wyns, S. Oscarson, H. De Greve, S. Hultgren, J. Bouckaert, *PLoS ONE* **2008**, *3*, e2040.
[10] a) N. Firon, S. Ashkenazi, D. Mirelman, I. Ofek, N. Sharon, *Infect. Immun.* **1987**, *55*, 472–476; b) T. K. Lindhorst, S. Kötter, J. Kubisch, U. Krallmann-Wenzel, S. Ehlers, V. Kren, *Eur. J. Org. Chem.* **1998**, 1669–1674; c) O. Sperling, A. Fuchs, T. K. Lindhorst, *Org. Biomol. Chem.* **2006**, *4*, 3913–3922; d) Z. Han, J. S. Pinker, B. Ford, R. Obermann, W. Nolan, S. A. Wildman, D. Hobbs, T. Ellenberger, C. K. Cusumano, S. J. Hultgren, J. W. Janetka, *J. Med. Chem.* **2010**, *53*, 4779–4792; e) T. Klein, D. Abgottspon, M. Wittwer, S. Rabbani, J. Herold, X. Jiang, S. Kleeb, C. Lüthi, M. Scharenberg, J. Bezençon, E. Gubler, L. Pang, M. Smieško, B. Cutting, O. Schwardt, B. Ernst, *J. Med. Chem.* **2010**, *53*, 8627–8641; f) O. Schwardt, S. Rabbani, M. Hartmann, D. Abgottspon, M. Wittwer, S. Kleeb, A. Zalewski, M. Smieško, B. Cutting, B. Ernst, *Bioorg. Med. Chem.* **2011**, *19*, 6454–6473; g) C. K. Cusumano, J. S. Pinkner, Z. Han, S. E. Greene, B. A. Ford, J. R. Crowley, J. P. Henderson, J. W. Janetka, S. J. Hultgren, *Sci. Transl. Med.* **2011**, *3*, 109ra115; h) J. Berglund, J. Bouckaert, H. De Greve, S. Knight, *Anti-Adhesive Compounds to Prevent and Treat Bacterial Infections*, Intl. Pat. PCT/US 2005/089733, **2005**.
[11] Glide, version 5.7, Schrödinger, LLC, New York, NY (USA), **2011**.
[12] I. L. Scott, R. V. Market, R. J. DeOrazio, H. Meckler, T. P. Kogan, *Carbohydr. Res.* **1999**, *317*, 210–216.
[13] C. A. Ocasio, T. S. Scanlan, *Bioorg. Med. Chem.* **2008**, *16*, 762–770.
[14] M. Prieto, E. Zurita, E. Rosa, L. Muñoz, P. Lloyd-Williams, E. Giralt, *J. Org. Chem.* **2004**, *69*, 6812–6820.
[15] M. Ishikawa, Y. Hashimoto, *J. Med. Chem.* **2011**, *54*, 1539–1554.
[16] S. Rabbani, X. Jiang, O. Schwardt, B. Ernst, *Anal. Biochem.* **2010**, *407*, 188–195.

[17] M. Scharenberg, D. Abgottspon, E. Cicek, X. Jiang, O. Schwardt, S. Rabbani, B. Ernst, *Assay Drug Dev. Technol.* **2011**, *9*, 455 – 464.
[18] J. Bouckaert, J. Mackenzie, J. L. de Paz, B. Chipwaza, D. Choudhury, A. Zavialov, K. Mannerstedt, J. Anderson, D. Pierard, L. Wyns, P. H. Seeberger, S. Oscarson, H. De Greve, S. D. Knight, *Mol. Microbiol.* **2006**, *61*, 1556 – 1568.
[19] G. Zhou, W. J. Mo, P. Sebbel, G. W. Min, T. A. Neubert, R. Glockshuber, *J. Cell Sci.* **2001**, *114*, 4095 – 4103.
[20] P. Aprikian, V. Tchesnokova, B. Kidd, O. Yakovenko, V. Yarov-Yarovoy, E. Trinchina, V. Vogel, W. Thomas, E. Sokurenko, *J. Biol. Chem.* **2007**, *282*, 23437 – 23446.
[21] I. Le Trong, P. Aprikian, B. A. Kidd, M. Forero-Shelton, V. Tchesnokova, P. Rajagopal, V. Rodriguez, G. Interlandi, R. Klevit, V. Vogel, R. E. Stenkamp, E. V. Sokurenko, W. E. Thomas, *Cell* **2010**, *141*, 645 – 655.
[22] D. Abgottspon, G. Rölli, L. Hosch, A. Steinhuber, X. Jiang, O. Schwardt, B. Cutting, M. Smieško, U. Jenal, B. Ernst, A. Trampuz, *J. Microbiol. Methods* **2010**, *82*, 249 – 255.
[23] a) J. E. Ladbury, *Biochem. Soc. Trans.* **2010**, *38*, 888 – 893; b) G. A. Holdgate, W. H. Ward, *Drug Discovery Today* **2005**, *10*, 1543 – 1550; c) R. Perozzo, G. Folkers, L. Scapozza, *J. Recept. Signal Transduction Res.* **2004**, *24*, 1 – 52; d) J. E. Ladbury, G. Klebe, E. Freire, *Nat. Rev. Drug Discovery* **2010**, *9*, 23 – 27; e) K. P. Murphy, D. Xie, K. S. Thompson, L. M. Amzel, E. Freire, *Proteins* **1994**, *18*, 63 – 67.
[24] a) B. Baum, L. Muley, M. Smolinski, A. Heine, D. Hangauer, G. Klebe, *J. Mol. Biol.* **2010**, *397*, 1042 – 1054; b) M. C. Chervenak, E. J. Toone, *J. Am. Chem. Soc.* **1994**, *116*, 10533 – 10539; c) J. E. DeLorbe, J. H. Clements, M. G. Teresk, A. P. Benfield, H. R. Plake, L. E. Millspaugh, S. F. Martin, *J. Am. Chem. Soc.* **2009**, *131*, 16758 – 16770.
[25] S. Cabani, P. Gianni, V. Mollica, L. Lepori, *J. Solution Chem.* **1981**, *10*, 563 – 595.
[26] A. V. Finkelstein, J. Janin, *Protein Eng.* **1989**, *3*, 1 – 3.
[27] K. P. Murphy, *Biophys. Chem.* **1994**, *51*, 311 – 326.
[28] E. Freire, *Drug Discovery Today* **2008**, *13*, 869 – 874.
[29] a) T. S. G. Olsson, M. A. Williams, W. R. Pitt, J. E. Ladbury, *J. Mol. Biol.* **2008**, *384*, 1002 – 1017; b) C. Diehl, O. Engstrom, T. Delaine, M. Hakansson, S. Genheden, K. Modig, H. Leffler, U. Ryde, U. J. Nilsson, M. Akke, *J. Am. Chem. Soc.* **2010**, *132*, 14577 – 14589.
[30] Phase, version 3.3, Schrödinger, LLC, New York, NY (USA), **2011**.
[31] a) B. A. Williams, M. C. Chervenak, E. J. Toone, *J. Biol. Chem.* **1992**, *267*, 22907 – 22911; b) E. J. Toone, *Curr. Opin. Struc. Biol.* **1994**, *4*, 719 – 728; c) T. K. Dam, C. F. Brewer, *Chem. Rev.* **2002**, *102*, 387 – 429; d) M. Ambrosi, N. R. Cameron, B. G. Davis, *Org. Biomol. Chem.* **2005**, *3*, 1593 – 1608; e) E. Garcia-Hernandez, R. A. Zubillaga, E. A. Chavelas-Adame, E. Vazquez-Contreras, A. Rojo-Dominguez, M. Costas, *Protein Sci.* **2003**, *12*, 135 – 142.
[32] H. Van de Waterbeemd, D. A. Smith, K. Beaumont, D. K. Walker, *J. Med. Chem.* **2001**, *44*, 1313 – 1333.
[33] M. Kansy, F. Senner, K. Gubernator, *J. Med. Chem.* **1998**, *41*, 1007 – 1010.
[34] A. Avdeef, S. Bendels, L. Di, B. Faller, M. Kansy, K. Sugano, Y. Yamauchi, *J. Pharm. Sci.* **2007**, *96*, 2893 – 2909.
[35] C. A. Lipinski, *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235 – 249.
[36] J. Kasuga, M. Ishikawa, M. Yonehara, M. Makishima, Y. Hashimoto, H. Miyachi, *Bioorg. Med. Chem.* **2010**, *18*, 7164 – 7173.
[37] R. Roy, S. K. Das, F. Santoyo-González, F. Hernández-Mateo, T. K. Dam, C. F. Brewer, *Chem. Eur. J.* **2000**, *6*, 1757 – 1762.
[38] a) F. Bitsch, R. Aichholz, J. Kallen, S. Geisse, B. Fournier, J. M. Schlaeppi, *Anal. Biochem.* **2003**, *323*, 139 – 149; b) S. Mesch, K. Lemme, H. Koliwer-Brandl, D. S. Strasser, O. Schwardt, S. Kelm, B. Ernst, *Carbohydr. Res.* **2010**, *345*, 1348 – 1359.
[39] G. Kemmer, S. Keller, *Nat. Protoc.* **2010**, *5*, 267 – 281.
[40] O. O. Krylova, N. Jahnke, S. Keller, *Biophys. Chem.* **2010**, *150*, 105 – 111.
[41] A. Ziegler, J. Seelig, *Biophys. J.* **2004**, *86*, 254 – 263.
[42] P. Artursson, J. Karlsson, *Biochem. Biophys. Res. Commun.* **1991**, *175*, 880 – 885.
[43] a) VCCLAB, Virtual Computational Chemistry Laboratory, 2005, http://www.vcclab.org (accessed May 3, 2012); b) I. V. Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V. A. Palyulin, E. V. Radchenko, N. S. Zefirov, A. S. Makarenko, V. Y. Tanchuk, V. V. Prokopenko, *J. Comput. Aid. Mol. Des.* **2005**, *19*, 453 – 463.

Received: March 6, 2012
Revised: April 27, 2012
Published online on May 29, 2012

